Development of a cost-predicting model for construction projects in Ghana

by

GEORGE HARRISON COFFIE

A thesis submitted in fulfilment of the requirements for the degree

Doctor of Philosophy

in

Engineering Management

Faculty of Engineering and the Built Environment

UNIVERSITY OF JOHANNESBURG

Supervisor: Prof. C.O. Aigbavboa

2018

Declaration

By submitting this dissertation for the qualification of Doctor of Philosophy in Construction Management at the University of Johannesburg, I declare that the entirety of the work contained therein is my own original work and that I am the authorship owner thereof (unless to the extent explicitly stated) and I have not previously submitted it in its entirety or part for obtaining any qualification at another university. I further declare that all sources cited or quoted are indicated and acknowledged by means of a list of references. I further cede copyright of this dissertation to the University of Johannesburg.

Signed:

Date:

Abstract

One of the foremost challenges faced by the construction industry is the issue of cost overruns. Cost overruns cut across construction projects of nations and continents as well. They vary in magnitude and occur irrespective of project size and location. Over the years numerous attempts have been made in the area of estimating cost of construction projects right and improving the efficacy or accuracy of cost estimating using different statistical methods.

This research investigated the factors that contribute to cost overruns and developed a predicting cost-estimating model for public sector building projects. The aim primarily was to extract factors from historical data of completed projects and use these predictive factors to develop a predictive model. Two models were developed using the predictive variables from historical data by the use of multiple linear regression and extreme learning machine. These models were compared to see the accuracy of performance.

Results from the study reveal findings that; predictive variables from historical data can be used to predict the cost of completion of construction projects at the contract award stage, the multiple linear regression model results as compared to extreme learning machine results shows that extreme learning machine performs better. The study brought to light the use of extreme learning machine for developing predicting cost-estimating models built on historical data from completed projects. This rarely exists in construction industry. It further substantiates the superior performance of extreme learning machine to multiple linear regressions using big data. The developed model can also be converted to desktop software for predicting completion cost by industry.

The contribution of this research is novel since the use of extreme learning resulting in an improved model for cost estimation and cost overrun needs more improvement and is limited in the body of knowledge.

Keywords: Construction cost, Predictive model, Regression, Extreme learning machine.

Acknowledgements

To God be the Glory, for His light is our light and our salvation: who shall we fear! All efforts would have been meaningless without His grace. Surely those that wait upon Him shall mount on wings like the eagle. In His own time He makes all things beautiful.

As the saying goes, ‘One tree does not make a forest’. This work could not have been successful without the help of several people who in their various way contributed to the output of this undertaking knowingly or unknowingly. The ideas and experiences they have shared affected my thinking and the development of this work. I owe them my sincerest thanks. I would like to express my sincere gratitude to my supervisor, Associate Professor Clinton Aigbavboa for their supervision, guidance and invaluable suggestions throughout the study. I am also indebted to my boss and Vice-Chancellor Prof. Emmanuel Kojo Sakyi of Ho Technical University for his support and constant checking on my progress and insisting that I work hard to complete the study timeously. I am thankful to my friends, Mr. Eric Adzivor and Mr. Emmanuel Afetorgbor, for their moral support, limitless patience and encouragement throughout my study. To my family and especially Miss Shiella Coffie, I say ayeekoo! These acknowledgments will not be complete without mentioning the contribution of Mr. Senyo Kofi Cudjoe for the invaluable assistance in my data analysis. To Dr. Bernard M. Arthur-Aidoo, I say thank you and know that you are a brother.

Finally, I would like to thank all the people who took the time during their busy schedules to honour my interviews and provide the data used in my study.

Dedication

This research is dedicated to my mother, Elizabeth Adzo Setrana, who struggled to take care of me from infancy; my wife, Seyram Gifty Delanya, and my children, Enam, Eyram and Elorm who endured the absence of a father throughout my studies. Without their understanding and tolerance this work would not have been a success.

Table of Contents

LIST OF TABLE

LIST OF FIGURE

Chapter 1: Introduction

1.1 Background

This research investigated the cost estimation of construction projects in the public sector of developing nations, with Ghana as a case study. Chapter I contains the background information relating to the study. This includes the significance of the study, the statement of the research problem, objectives of the research, research questions, and the research design and approach. It further seeks to establish the background information of cost estimating in the construction industry.

1.2 Cost estimation and construction cost overruns

As the Biblical question goes, ‘’ “Will anyone of you embark on building a house without knowing the cost first?’’ All construction clients have one main concern, namely to construct a facility that is functional, at the required quality and delivered within acceptable budget and with an expected reasonable duration. This can be achieved through a cost-estimating process, among others.

According to Peter (2005), estimating is the process of determining the expected quantities and cost of the materials, labour and equipment for a construction project and this must be done as accurately as possible. This process plays a key role in the operation of organizations. It does not only lead to a successful bid and profitable margins but must also be seen to guide and influence the way projects must be delivered.

Hence to estimate accurately, one must have a sound understanding of the construction methods and materials, as well as the capacity of skilled labour. Furthermore, the estimator requires the skills to determine the quantities of materials, labour and equipment needed to complete the job. Further, updated knowledge of the prevailing market prices of these determinants is necessary. Many factors influence the quality of a construction budget and the estimating process: the method of cost estimating that is chosen is one of these that should be taken seriously. It thus serves as part of the compass for navigating through the execution and completion of any construction project.

The art of establishing the cost of infrastructure in the construction industry has gained much interest among construction experts with the aim of procuring a facility that meets, among others, the expected budget. The accuracy of estimating the cost of construction project is a critical factor in the success of the project (Kim, An & Kang, 2004; Ling & Boo,2001; Laryea, 2010), yet this remains uncertain in the construction industry (Akintoye, 2000; Akintoye & Fitzgerald, 2000; Laryea, 2010). Earlier researchers such as Skitmore & Wilcock (1994), Aibinu & Pasco (2008), Ling & Boo (2001) and Doloi (2011) concurred that estimating activities are considered questionable, outdated and difficult. If this is true, there is a need to re-examine the process to reveal what extra factors account for the uncertainty.

Previous research on cost overruns of construction projects has established that the issue of construction projects exceeding their budgeted cost remains unresolved and seems to be on the increase rather than decreasing. Flyvbjerg, Mette, Skamris Holm & Buhl, et al. (2002), in his seminal works concludes that nine out of ten construction projects experienced cost overrun on their budget. According to Flyvbjerg et al., (2002, 2004), the probability of infrastructure project experiencing cost overrun is 86% and the average size of overrun can be as high as 45% for rail projects, 34% for bridges and tunnels and 20% for road projects. Other reports for the building sector indicate that overruns in some cases exceed the initial budget and original cost by 70% (Love, Edwards, & Irani, (2012) and by 183% for highway projects (Odeck, 2004).

It can be confirmed that cost overruns cut across all projects irrespective of types: houses, roads, railways, bridges, tunnels and schools (Hinze, Selstead, & Mahoney, (1992); Love, Smith, Simpson, Regan, Sutrisna, & Olatunji, 2014). The size of project does not matter, whether small or mega, they all experience cost overrun (Bordat, McCullouch, Sinha, & Labi, 2014). In a study by Flyvbjerg et al. (2002), 285 projects had cost values ranging from small ($1.5 million) to mega ($8.5 billion). His study of cost overrun on five continents reveals that overruns occur across almost all the countries, whether developed or developing.

What is of significant is that Flyvbjerg et al. (2002) observed in their studies that the size of overruns has not improved in seventy (70) years. The ever-increasing trend continues till the present day. He concludes arguably that no learning that would improve cost estimate accuracy seems to take place. It is in this regard that this research employed an extreme learning machine to develop an improved model to predict the construction cost of projects. The study also endeavours to answer questions such as what accounts for the lack of improvement of the accuracy of cost estimates in the construction industry; what can be done to improve the closeness of estimates to actual costs and whether this level of accuracy can be achieved.

Since achieving a reliable cost estimate is elusive, the question can also be asked whether a review at stages in the project life cycle would to help to improve the reliability of cost estimates. There has been some improvement in the way projects are managed in modern times. However, the evolution of the procurement system has brought about a change in project management from the traditional methods to more hybrid methods. With the advancement in information and communication technology (ICT), various tools and techniques have been developed to enhance the delivery of projects. Considerable research has been undertaken in the area of project management theory and practice. The advancement of construction cost estimating models is on the increase: some are based on regression analysis, neural networks and case-based reasoning. Information technology in the field has improved tremendously. Computer-aided design has also given rise to 3D and 4D modelling, leading to improvements in managing projects. Software development in IT has been on the increase, thereby enabling statistical modelling to be used in the estimating of the cost of projects.

Therefore, this research seeks to evaluate the problem of cost overruns, initiate a review of its roots, determine the magnitude across the phases of the project life cycle and provide some possible solutions to the problem.

1.3 Research problem statement

Cost overruns have become a constant feature of infrastructure projects. The success of these projects is measured based on specific factors of which cost is one. The accuracy of estimating construction cost is a critical success factor in project delivery (Kim, Seo, & Kang, 2005). The availability of improved cost estimation techniques will facilitate effective control of cost in construction projects. Hegazy (2002) suggest that cost estimating is neither simple nor straightforward owing to a lack of information. Even though models developed were appropriate for estimating cost, there are limitations in their usage and relevance. Kim, An & Kang (2005) advocate for further research to develop a hybrid model that will integrate the various tools. Hence, the aim of the current research is to develop an improved model using an extreme learning machine.

There are contradicting figures as to the margin of cost overruns while Flyvbjerg et al., (2004) suggested that average cost overruns for infrastructure projects can range between 20.4% for roads, 33.8% for bridges and tunnels, and 44.7% for railways. Love, Edward & Irani (2012) reveal significantly lower levels of average cost overruns, with 13% for roads, and 5.5% for bridges. According to Love, Edwards & Irani (20012), checks and verifications need to be undertaken throughout each phase of the design process to prevent errors. However, the mixed results from researches demand a review of the cost estimating processes with a special look at the project life cycle.

A thorough review of the factors that cause cost overruns and the stages of the project life cycle that these causes affect most must be conducted. To achieve this, it is necessary to examine the project life cycle from the inception stage to the award stage and from the award stage to completion stage.

One of the most notable construction industry clients in Ghana is the government. There are various ministries and agencies that work in collaboration and are charged with the construction and the maintenance of government or public infrastructure throughout Ghana. The role of these entities exceeds setting the scope and financing the project. Their activities usually span from the inception stage to the project close-out stage. These also involve the estimation of project cost.

To this team, it is important that accurate and reasonable construction estimates be established to the extent of forecasting the likely cost overruns. Several efforts have been made to forecast the accurate cost of construction projects. Regardless of these efforts and methods adopted, these estimates at inception are frequently not close to the actual cost and the real cost overruns emerge more often than not when projects are completed and all details are known.

Furthermore, there is no consensus regarding the factors that influence cost overruns. Researchers have indicated that beyond the models or methods used, the experience of the personnel also plays a key role in the accuracy of the project cost estimate. However, there has not been much research to determine the statistical significance of these factors. The capability to predict the future cost based on past data is an important tool to support construction clients in decision-making. Multi-step ahead forecasts are needed for tactical evaluations, such as planning production resources. The past can be consulted in the form of using data on completed projects to predict future cost of projects early stages of project delivery Improvements in information technologies have created the possibility of collecting and processing considerable datasets. Data mining (DM) techniques aim at extracting useful knowledge from raw data. There are several DM algorithms, each one with its own advantages. If the estimating process is not straightforward and simple, then digging deeper to mine data is important in order to unravel the latent factors that actually cost overruns and to improve the predicting of cost.

Further research is required to quantify and model the relationships that exist among the factors that cause cost overruns, the interaction between these factors in order to develop models that can predict construction cost overrun when given the conceptual level of the project data.

There is a need to develop a predicting cost-estimating model to forecast the likely final cost of the project based on the data of influential factors that contribute to cost the overrun of construction projects. Therefore, the problem statement for thesis is the following: Given that no adequate predicting method or process of determining cost estimating at tender or inception stage of construction project are known, the current thesis will contribute to this body of knowledge. A cost estimating model is developed in this thesis.

1.4 Aims of the study

The aim of this research is to examine the cause of cost overruns in the public sectors with a view to developing a predictive cost-estimating model.

1.5 Motivation for the study

Cost is one of the most important factors, if not the most important, when procuring any infrastructure project. The motivation for this study was to develop a predicting cost-estimating model that would predict a reviewed probable final cost of projects and point at the least possible cost overrun. This thesis studied the determinants that most significantly influence the final cost of projects. These determinants were then integrated into the model that was developed. The existing relationship between these variables and the newly-found gaps were considered in developing the model while taking their level of importance into consideration.

1.6 Significance of the study

1.7 The study

This study contributes to the body of knowledge by synchronising relevant literature that supports the investigation and establishing the factors that contribute to the final cost of construction projects. It revealed the most influential factors that contribute to construction cost overrun. The information from this exercise and the developed model would enhance the cost-estimating process and inform public sector clients of the importance of proper budget conducting reviews on construction cost before projects are awarded.

1.7.1 Research questions of the study

This research answered the following broad questions:

- Can factors influencing construction cost overrun be identified through an analysis of historical conceptual level project data and statistical algorithm analysis carried out with them to develop the final cost of projects?
- What are the relationships between the identified factors influencing cost overrun-influential factors and the public sector construction cost overruns?
- How do these factors interact to describe and model the construction cost given the conceptual level project data?
- Can the Extreme learning machine technique outperform the multiple linear regression technique (MLR) already in use in forecasting construction cost?

1.7.2 Research objectives

This thesis studied of construction cost overrun in the construction industry of Ghana. In order to achieve this goal, the following objectives were formulated:

- To determine appropriate, cost overrun influential factors of construction projects and apply them to develop a construction cost prediction model in the construction industry of Ghana;
- To investigate the pattern and trends underlying cost overrun in the construction industry of Ghana;
- To identify factors that affect cost overrun in the construction industry in Ghana and select those that are statistically significant;
- To apply the extreme learning machines (ELM) and multiple linear regression (MLR) technique to develop a cost-estimating model using the selected significant factors;
- To apply the MLR method to develop a forecasting model to serve as a benchmark to judge the performance of the extreme machine learning model;

1.7.3 Research methodology and design

A mixture of both qualitative and quantitative methods of research was employed to extract information and collect data for decision-making. This was done to mitigate the deficiencies in both methods. A case study approach based upon analytic induction was used to further examine the underlying dynamics that contribute to cost overruns in public sector projects. This was exploratory in nature, based on interviews with personnel and data from completed project sources. The research was conducted in Ghana and interviews were conducted within public and private organizations. These project data involved were in relation to educational projects only. Consultants, both public and private, who engage in government infrastructure projects and contractors were interviewed. The personnel cut across the following places of work: The Ghana Institution of Engineers, Ghana Institution of Architects, Ghana Institution of Surveyors, private consultants who engage in government infrastructure delivery, Ministries, Metropolitan, Municipal and District assemblies MMDAs), government agencies and other related entities and contractors who execute government contracts in Ministry of Education infrastructure delivery. The selection of the personnel involved in the interviews was based on the knowledge of and work experience in government infrastructure project management.

1.7.4 Data collection

According to Lapan & Quartaroli (2009), any data collection method must fit the purpose, respond to the needs of the research and also suit the time frame, which is allocated to the project.

There are several ways of collecting data. The four main ways as suggested by Lowe, Thirpe & Easterby-Smith (2004) include interviews, questionnaires, test or measure and observation. In this research explorative interviews (unstructured) were conducted after a thorough review of literature. Multiple linear regression and extreme learning machines were used to develop models from which the improved model was determined.

The methodology in this research followed the following steps:

- Establish the need for the research by understanding the current cost overrun issues and cost-estimating practices and procedures employed by the construction industry;
- Identify and select a construction cost and cost overrun information data source;
- Collect and compile cost-related data;
- Seasonally adjust the contract cost and cost overruns collected to calculate a historical and typical construction cost;
- Conduct statistical regression analyses to identify cost overrun-influential project factors and develop models for predicting construction cost;
- Validate the models developed;
- Describe the proper use and bounds of the models developed; and
- Discuss the applicability of the research methodology to public sector construction and the construction industry in general.

Therefore, multiple linear regressions and an extreme learning machine was used to develop the model.

1.7.5 Delimitation of the study

The scope of this study was limited to cost overruns and the cost of public sector construction projects in Ghana. It focused on the significant factors that influence the cost overruns of construction projects. Gaps in literature regarding these factors were identified and evaluated for the development of a mathematical cost-estimating model for construction projects cost overrun.

1.7.6 Limitations of the study

The scope of the study was limited to clients, consultants and contractors in Ghana since these groups form the major players in the construction industry. The targeted group of contractors and consultants was based on their classification according to Ministry of Water Resources, Works, and Housing in Ghana who undertakes publically funded projects. The limitation was necessary to allow homogeneous data to be collected for the study, taking into account the limited time and logistics available to undertake the research. To achieve the purpose of this study within time, logistical and financial constraints, a complete national survey was not conducted. The findings and proposed model, however, may be generalized since the structure, requirements and implementation of projects in the public sector in Ghana follow a standard form.

1.8 Ethical considerations

All ethics were observed such as awareness of the sensitive nature of government businesses in this democratic dispensation. In addition, respondents were assured of their privacy and the confidentiality of their responses. Moreover, participation in the research was entirely voluntary.

1.9 Structure of the study

The thesis comprises seven (7) chapters. The contents of and the linkage between chapters are outlined below:

The first chapter provides an overview of the thesis. The background, problem statement, aims and objectives, motivation, significance, research questions, methodology and design, delimitations, and structure of the research are addressed here.

The second chapter consists of a comprehensive literature review of cost overruns in the construction industry. The magnitude of cost overrun problems are reviewed, sources of cost overruns are discussed, and theoretical concepts of cost overruns, cost estimating models and factors affecting cost overruns are reviewed as well.

The third chapter discusses the construction industry, cost overrun issues and matters related to the case of Ghana, and finally, cost overruns issues in Nigeria and South Africa.

The fourth chapter focuses mainly on multiple regressions using the statistical selection of factors influencing construction cost overruns. Data from completed projects are extracted. Significant factors influencing cost overruns arising are selected through the STEPWISE procedure.

The fifth chapter reviews literature on the theory and the application of the ELM. A justification of using the ELM to forecast the cost is highlighted and how the model was developed is discussed in the chapter as well.

The sixth chapter presents the results and analysis of three forecasting models using alternative forecasting approaches. The first model employed the multiple linear regression (MLR) technique using the statistically significant indicators selected. The second model used the ELM technique and, similarly, used the significant factors used for the first model. The third model used the same ELM technique but made use of all factors identified. The models were compared and discussed.

The seventh chapter presents a summary of the findings, and the conclusion. In addition, recommendations are made and the contribution to and limitations of the study as well as suggestions for future research are outlined.

2.0 Conclusion

Chapter 1 introduced the theme of the research under study. It gave insight into the research structure, the background of the study and its significance. It communicates information on how the research report is presented. The next chapter examines the theoretical and contextual perspectives of construction cost, cost overruns and other relevant literature related to the objectives of the study.

Chapter 2: Early Construction Project Cost Forecasting

2.1 Introduction

This chapter presents a detailed discussion on the need for early cost estimating and the problem of construction cost overruns of projects in the construction industry. In addition, it gives an overview of the magnitude of cost overruns and its related issues. It also presents the theoretical explanations for the causes of cost overruns. Finally, a more detailed trend cost-forecasting models development is examined in this chapter.

Cost, quality and time are the three cardinal issues that concern construction clients' worry who want to know how their projects will fair within these contexts. Clients want projects to be delivered to acceptable budget, of the best quality and within an acceptable time frame. The cost of projects is of great importance to clients at the early stage as it aids them in making informed decisions. Thus effective cost planning of building must take account of the cost of the project which needs to be within an acceptable limit of budget whilst not playing down on quality, risk, time, probable changes in scope, aesthetics or utility consumed.

The art of getting the cost right from the inception is critical to the success of the project and it has far-reaching effects on the rest of the project delivery stages. Estimating methods of various kinds are used in the preliminary stages of the inception of projects to forecast the cost of projects. This is done with limited available project information and is available to project managers enabling them to come up with more reliable control of costs in construction projects. The importance of these models cannot be overlooked; however, their usage or application is neither simple nor straightforward. The estimating process must take into account all possible events that will occur throughout the project life cycle. Downplaying or ignoring this fact will set the project delivery back.

2.1.1 Getting estimates right

Cost planning can be done effectively by projecting the cost of building design taking into account quality, risk, possible determinants of cost and cost changing factors. It is done to ensure that the project is going to be within the budgeted limit of clients. The project development process comes with stages: salient decisions are made during the conception stages and these decisions have economic effects that are significant at the later stages. Estimates are made to project the likely total cost: however, the art of estimating is nether straight neither forward nor simple. Accuracy of estimates in this regard is of much economic importance and they must take into account all possible future occurrences that would likely influence the cost of projects.

The cost of components such as materials, labour, transportation, profit and overheads are not the only factors that need attention: other factors such as the type of client, type of project, type of contract, possible design and scope changes, site conditions, project duration and tendering method must be considered as well. The event of accounting for all these factors in terms of cost is challenging while overlooking them would pose serious cost overrun issues.

The demand for enhanced preliminary construction project cost appraisals was postulated by various writers and much investigation in this sphere has been undertaken by construction industries across the globe, thereby aiding the improvement of these practices (Bromilow, 1980; Nkado, 1992; Chan & Kumaraswamy, 1995; Chan, 1999; Skitmore & Ng, 2000; Burrows, Pegg, and Martin, 2005). Unfortunately, many research studies appear to lack the skill of accuracy and how to use historical data to predict the cost of projects at the initial stages of the project. Forecasting construction project cost has occurred in the construction industry globally (Bromilow, 1980; Nkado, 1992; Chan & Kumaraswamy, 1995; Chan, 1999; Skitmore & Ng, 2001: Burrows, Pegg, and Martin, 2005). Nevertheless, there has not been much evidence of this in the domestic public sector in Ghana. The next section highlights the extent of cost overrun and its associated issues in construction.

2.2 The level of cost overrun and its associated issues in construction

Research has provided overwhelming evidence to substantiate the fact that cost overrun has been an issue of concern in the construction industry. However, what pertains to cost overruns in Ghana and why they are so prevalent in the construction industry must be discussed as well.

Globally cost overruns are experienced in the construction industry from the smallest to largest contract, whether complex or not. Developed, developing and underdeveloped countries on all continents contend with the issue of cost overrun in the domain of infrastructure delivery. The evidence of these phenomena is overwhelming. In Massachusetts, United States of America (USA), it is on record that close to 50% of bridge and road projects in construction overrun their budgets and 33% were completed beyond the estimated time (Love, Sing, Wang, Irani & Thwala, 2014).

Notable projects like the Wembley Stadium in the United Kingdom (UK) had an associated 50% cost of overruns; the parliament building project in Scotland exceeded its budget by more than 900%; the Perth Arena project which had a contract value of 168 million (AUD) overran its original value by more than 300% (Love, Edwards & Iran, 2011). Flyvbjerg (2009) and Odeck (2004) suggest that across five (5) continents the issue of cost overrun is substantial and this issues spans over seven (7) decades.

Cost overruns are so prevalent in the construction industry and cut across works such as bridge construction, rail construction, road construction and building construction. However, they vary in magnitude along these works. (Love et al., 2012 and Flyvbjerg et al 2005).

The Egan Report 1998 (which was commissioned by the UK government) on the performance of the construction industry indicated that more than 50% of projects exceeded budget estimates. The General Accounting office in 1997 in the USA reports that close to 77% of projects overspent their budgets, even by as much as 200%.

In the city of Edinburg Council 2014 report the Trams project had forecast a cost of £375 million (ZAR 6,225,982,405.50) but was completed late (three years in excess of the duration) at the cost of £776 million (ZAR 12, 884, 811,515.58). Whatever the range of cost overrun is, clients in the industry are concerned and the root causes must be ascertained and proper measures formulated to curb them.

2.3 Cost overrun: The root causes

Globally, construction projects' contracts are being entered into daily. Regardless of the type of contract, size, construction, whether simple or complex and even the kind of clients and the other parties involved, there is one issue that is common: they can all experience cost overrun. Several varying but growing schools of thoughts has been suggested on the causes of cost overrun. Research works suggest in one vein that the strategies by practitioners lead to cost overrun, that is to say they intentional give estimates that are not realistic. (Flyvbjerg et al., 2002,2004, 2009; Wachs 1990, 1989 and Lavallo & Kahneman 2003). The other schools of thought suggest that it not intention that practioners do not get estimates right rather they get it wrong as a result of uncertainties. (Akintoye & Maclead 1997; creedy 2006; Barccarini 2005, Kahneman & Tversty 1979 and Kahneman 2001). The last schools of thought put the causes of cost overrun on the basis that projects evolve over time and there are events and pathogens that emerge along the delivery phases of projects. Pathogens like rework among others lead to cost overruns. (Love et al. 2012, 2014; Osland & Strand 2010, Odeck 2004 and Love et al. 2015). This study is constructed along the last view. This work does not intend to replicate these roots but rather takes a look at how they are identified and classified as well as their relatedness to the project life cycle. This can be achieved by dwelling on the related works done on the causes of cost overrun by scholars in and developing predicting models using the causes of cost overrun.

2.3.1 Risk

Ahiaga-Dagbui (2014) suggests the word 'risk' has been misused, describing it as uncertainty of any kind without guidance as far as the contingency of projects is concerned but later insists that uncertainty must be taken in a sense radically distinct from the familiar notion of risk. This suggests that the use of the words ‘uncertainty’ and ‘risk’ interchangeably do not connote the same meaning. Lately, Ross & Williams (2013) referred to risk as the threat or possibility that an action or event will adversely or beneficially affect an organisation's ability to achieve its objective. They expand this explanation this by suggesting that word ‘risk’ can be described as the consequence of a hazard, measured as the likelihood of the hazard and its severity, in the event of the hazard occurring. A situation where possibility or prediction of an event occurring is not known is referred to as ‘uncertainty’. In this work, the primary concern is construction cost, thus risk definition is limited to the modest definition as the measure of manifestation and likelihood of financial loss occurring during project delivery.

Construction projects by their nature are highly disposed to the effects of risk and uncertainty they are distinctive, multifaceted and changing; there are several key players associated with handling each project with conflicting project and business objectives. Construction projects experience different weather, unpredictable ground conditions (in most cases) and can span several years in their duration of execution. As such, risk has been frequently mentioned as one of the key causes of construction projects not meeting expected budget (Akintoye & MacLeod, 1997; Creedy, 2006). Debatably, the construction industry is possibly one of the industries that are most risk averse, with the cost of project being one of core areas vulnerable to its effects. Practically and interestingly all forms of risk, whether scope changes, inclement weather, unsuitable ground conditions, contractual arrangements, disputes, clients’ cash flow problems, among others, bring about certain financial difficulties.

Despite the fact that risk and uncertainty appear to permeate the construction industry, Baccarini (2005) and Burger (2003) suggest that mostly they are either disregarded or accounted for in a manner that is illogical: by merely allocating percentages popularly called contingency funds. The task of managing risk, according to Flanagan & Norman (1993), is frequently inadequately done, so that considerable risk is latently retained, eventually causing cost escalation during project delivery.

Conversely, is there a possibility of combining intuitive decision and the predicting of future events that can be clear-cut or fair-minded? Love et al. (2012) suggest there is not. They are the postulators of prospect theory whereby they considered how decisions are made under conditions of risk and uncertainty. Flyverbjerg et al., (2005) suggests that with minimal or vague information, decision makers tend to decide based on the potential gains, or losses, of a project, and not essentially on the actual outcome of the decision they make. In the Noble Prize-winning work of Kahneman (2011) on decision making and behavioural economics, he describes decision making and the illusion of understanding, asserting that we regularly display an excessive conviction in what we think we understand about any situation, and that our incapability to recognise the full extent of our ignorance and the uncertainty of the events in this world we live in makes us disposed to overestimating the depth to which we actually understand issues (Kahneman, 2011). We are usually inclined towards overlooking or misjudging the probability or gravity of likely risk events. The prospect theory (1979) can be extended to decision-making in the construction industry, specifically for complex public projects where the cost of risk and uncertainty are certainly sensitive. This theory tends to substantiate the contemporary seminal works of Flyvbjerg et al. (2002, 2004) on strategic misrepresentation and optimism bias.

2.3.2 Strategic misrepresentation

Seminal work by Flyvbjerg et al. (2002, 2004) and research by Wach (1989, 1990), suggests that one of the causes of cost overrun is strategic misrepresentation. They suggest that strategic misrepresentation, the deliberate or understating of the amount, time and other resources necessary to deliver the project, is probably the principal source of cost overrun, especially in the case of large public sector projects. Flyvbjerg et al. (2002) undertook a statistically significant study, analysing the cost performance of 258 transportation projects valued at US$90 billion and classified the sources of cost overruns on construction projects into four groups: technical (error), psychological, economical, and political. They compared the budgets of projects as at the time of the decision to build to the cost at completion. In their results there are discrepancies in cost forecasts for transportation infrastructure projects and these range from an average of 45% for rail, 34% for bridges and tunnels, and 20% for roads. In their sample, nine out of ten projects exceed their cost targets, indicating that infrastructure projects in their works have an 86% probability of exceeding their budget.

Flyverbjerg et al. (2012) observed that the issue of projects outrunning budgets were not randomly distributed but were systematic, thus concluding that the cost estimates used to seek client approval to construct were highly and systematically misleading. (Flyvbjerg et al., 2002), and attributed this to foul play by project sponsors. For a project to be approved, Flyvbjerg (2002) suggests that promoters and estimators working on public works tend to purposely underestimate the real cost of the project.

2.3.3 Optimism bias

Advance work on the strategic misrepresentation point of view by Flyvbjerg et al. (2002) postulate theories on optimism bias, stemming from the work of Weinstein (1980). Optimism bias is described as the disparity between a person's expectation of an event that is yet to happen and the outcome of the event as it happens. Thus if predicted outcomes are better than the actual outcome, the bias is optimistic; if the actual outcome is better than the predicted outcome, the bias is pessimistic. The degree of the optimism bias is therefore measured empirically by taking note of an individual's predictions before an event occurs and comparing those with the real outcomes that emerge. Drawing on this theory Flyvbjerg et al. (2004) submit that decision-making in policy and infrastructure planning is inconsistent in that planners think they know, or they are familiar with all conceivable series of project events from the point of decision to build to project completion. This results in planners having too much confidence in the prospects of projects and this leads to impracticable estimates. Optimism bias might, unlike the case of strategic misrepresentation, not be influenced by deception, but it frequently leads to miscalculating actual cost, exaggerating benefits, and not taking into account the likely effects of mistakes and uncertainty.

2.3.4 Works beyond strategic misrepresentation and optimism bias

Love, Edwards & Irani (2012) extended the work done by Flyvberg et al. (2004) on the postulation that there are two main reasons why projects experience cost overruns according to Flyverberg et al., (2004) namely strategic misrepresentation and optimism bias. It was explained that psychological explanations account for optimism bias and political stakeholders judge future events in a more positive light rather than applying the test of actual experience. Thus, managers and estimators deliberately and strategically exaggerate the briefs of their project. They further present unrealistic estimates by underestimating to gain approval and funding. Love et al. (2012) in their work concluded that a significant number of projects experience overruns without strategic misrepresentation or optimism bias being present. It is clear from their work that these two phenomena are not the only causes of cost overruns.

Love et al. further suggested that there are underlying pathogens resident in the project environment that play a key role in influencing cost overruns and these are different from strategic misrepresentation and optimism bias. This has been collaborated by Gil & Lundrigan (2012) in their work in which they carried out a comprehensive assessment of the reason for cost growth on construction projects. They were of the view that projects evolve; thus, construction projects change considerably between the conception stage, the inception stage and the completion stage and these changes give rise to cost overruns. By extension, Love et al. (2014) suggest that misrepresentation and optimism bias lack confirmable causality and can have limitations in their application during the estimating of cost overruns.

2.3.5 Organisation of project: Governance

From the technical report of the Centre for Infrastructure Development, University of Manchester (CID report No.3/2012), Gil & Lundrigan (2012) carried out research on the leadership and governance of projects, and advance the argument that there should be an all-inclusive assessment of cost growth. Love et al. (2014) support the same views. Their argument mainly dwells on the fact that all projects devolve. Most often, construction projects vary substantially in design and scope throughout the project life cycle, sometimes owing to changes directed by clients or else owing to technically required changes. This implies that it may be inaccurate to merely relate project cost at inception, X, with that of completion, Y, and assume that since Y >X, cost overruns have occurred and estimates at inception X were done deliberately to conceal reality or the incompetency of the estimators gave rise to X. The costs of X and Y will be different. A comprehensive justification of overruns needs to consider the process and the product as well as the sources of changes to scope. According to Love et al., (2014) and Gil et al. (2012), project overruns are not actually the case of projects not going according to plan (estimated cost), but the other way round – plans not going according to project.

A recommendation by Gil & Lundrigan (2012) introduces a "relay race" framework to enable a full analysis of cost growth, particularly on extra-large projects (mega) such as the London Olympics Park project, the Parliament of Scotland or Terminal 2 at Heathrow Airport, all of which appeared to have suffered the same fate of cost growth, at least from a cursory analysis. The suggested relay race framework of construction delivery explains that project leadership changes hands; thus, like a baton, it should be passed on from one person(s) or organisation at the various stages of the project life cycle. The features of projects and promoters (such as aims, scope, and the skills and proficiencies of the project sponsors and promoters) during the conceptual stage vary vastly from their counterparts' other stages of the project life cycle. Furthermore, it is a common phenomenon for most public projects to have long incubation periods, delaying for several years before final approval is given, by which time project budget would also have changed countless times and this also adds to several factors (events and actions) contributing to cost overrun of projects.

In the work of Odeck (2004), he suggested that the instability of the input cost of construction leads to uncertainty in predicting them financially and accounts for underestimating the cost. Factors such as the escalation of material prices, inflation, hitches in the procurement of construction materials at stable prices, unstable wages, increases in the cost of labour owing to environment restraints, regular payment problems from agencies, cash flow problems in construction, the financial stability of clients contractors, delayed payment for certified completed works, instability of currencies, excessive interest rate on loans, cash flow problems faced by contractors, and the unavailability of materials, among others, impact the input cost.

Shenhar & Dvir (2007) point out that the newness (novelty) of projects and their possible solutions come with vagueness of tasks and outcome, hence making planning and estimating challenging. Inherent factors (in the newness of projects) such as the inexperience of the project environment or location, a lack of expertise of project type, insufficient contractor expertise, unforeseen subsoil or ground conditions, technical incapability, unconstructable and complicated design, lack of contemporary equipment (technology), unforeseen weather conditions, difficulty in geological conditions, unpredicted site conditions, site limitations, and lack of standard method of quantity take-off are associated with the newness or novelty of projects.

The complexity of projects gives rise to difficulties in the coordination of tasks and the parties concerned, thus making it difficult to achieve set targets. Baccarini (1996) described project complexity as consisting of countless diverse interdependent parts and can be operationalized in relation to differentiation (the number of diverse elements, e.g. tasks, specialists, components) and interdependency (the degree of interrelatedness between these elements). The following are notable causes of cost overrun where complexities of projects prevail: Flaws in the infrastructure, problems associated with labour, insurance issues, work security challenges, health and safety problems, added works, contractors' site management and supervision problems, unavailability of site workers, deficiency in communication among parties, and errors during construction. In addition, there are imprecise site investigations, modifications in material specification and type, design errors, project magnitude, unfinished drawings, incompetent specifications, scarcity of skilled labour, equipment availability and failure, and the number of works being done at same time. Finally, a lack of constructability, scope changes of the project, insufficient equipment, labour disputes and strikes, owner interference, obstacles from government, laws and regulatory frameworks, delays in the preparation and approval of drawings, disputes on site, and political complexities can all lead to cost overrun.

Demands of the project, and insufficient time to complete the job i.e. time pressure (Shenhar & Dvir, 2007) compels project crews to take short-cuts or undertake work tasks concurrently causing delays and overrun. Similarly, impractical contract duration and requirements imposed, inappropriate planning and scheduling by contractors, interruption in material procurement, poor design and late design, bottlenecks in the delivery of materials and equipment, untimely decision making, insufficient planning and scheduling, not honouring on-time payment due to suppliers or subcontractors and inadequate time for estimates are likely to exist in the delivery of construction projects.

Love, Edwards, Smith, Irani (2012) and Walker (1995) assert that regardless of deceitful practice or unfounded optimism, infrastructure projects remain subjected to substantial levels of rework, which predominately arises through design changes and errors. Further works by Love et al. (2012) affirm that intermediary events and actions contribute to cost overruns.

Love et al. (2012) suggest that the method of procurement used in selecting the design consultants affects the cost of projects. In their work, it is mentioned that governments argue the traditional lump sum method of procurement whereas their hybrids offer cost and time certainty, lower risk and great clearness and answerability. However, Cheung,Yiu & Chim (2006) suggest traditional procurement methods do support the smooth management and delivery of projects. Raisbeck (2010) concluded that traditional procurement methods fail to meet the procurement triangle (time, cost and value) and are associated with dispute. But what is of significant is that Love et al. (2012) affirm that despite the procurement methods, design errors are not largely influenced by them.

In the findings of Love et al. (2012), activities of strategic misrepresentation and optimism bias could fuel pathogens to occur within the atmosphere in which projects are procured. Pathogens are referred to as latent conditions that remain inactive until an error is detected and designers do not usually know about these.

Pathogens contributing to occurrence of errors were further categorised by Busby & Huges (2004) as practice: resulting from deliberate practices by team player; Task: stemming from the nature of work to be performed; Circumstances: ensuing from the environment or situation within which the project is being conducted; Organization: organizational structure or operation gives rise to this; System: emanating from the system of the organization; Industry: ensuing from the structural property of industry; and Tool: emanating from the practical physiognomies of the tool.

2.4 Cost overrun specific factors

The specific factors enumerated by various works are countless; Okmen & Oztas (2010) assert specifically that cost growth will lead to overrun from risk and uncertainty. Love et al. (2012) examining contract documentation reveal that the specific factors are scope creep and rework. Lavallo & Kahneman (2003) mention optimism bias and the work of Wachs (1990) and Flyvbjerg et al. (2009) attest to foul play and corruption.

Love et al. (2012) reinforce the factor of rework and design changes. Odeck (2004) identified with project durations, project location, costs estimate at contract award and the size of projects as factors that influence overrun.

Love et al. (2012) explain that rework comes as a result of ineffective use of information technology, staff turnover or allocation to other projects, incomplete design at the stage of tender, insufficient time for contract preparation and poor coordination between design teams.

Bordat et al., (2004) suggest change order is the cause of overrun and is due mainly to errors and omissions in design while Gil & Lundrigan (2012) reveal that overrun comes as result of changes either proposed by the client or technically imposed.

2.4.1 Factors contributing to cost overrun

The list of factors that contribute to construction cost overruns is endless and, depending on whatever lens is used to view it, the classification can be numerous and to some extent overlapping. Love et al. (2012) suggest a fundamental shift from the use of competition by way of tendering in selecting design consultants. Love et al. (2012) further advocate for design review mandates, and cross checking to be undertaken throughout the phases of the design process for errors to be detected and reduced. However, it is necessary to move beyond this and identify the factors along the project delivery that lead to cost overruns. Projects go through life cycles (phases), as it is evident that causes of overruns are countless and they have similar characteristics of impact on the cost at completion or close out. It will be technically useful and conceptually meaningful to develop a classification of causes based on their impact on overruns and relating effects along predefined dimensions or at specific phases of the project. This will aid relay in project governance and mitigate the latent effect overrun can have on project cost.

Literature suggests that more research reveals that the scope of projects makes a significant contribution to the resultant cost of projects and the time it takes to complete them (Ng et al., 2001; Bromilow, 1969; Walker, 1995). Scope of projects refers to the measurement of its size in relation to function. Attributes such floor areas, storey height or the number of storeys and building volumes were influential in the scope of projects. Love et al. (2012), suggests that the larger and more complex a project is, the greater the possibility of an increase in cost overrun. This is as a result of errors in estimating and variation orders coming from the initial contract sum at award stage, changes in the number of storeys, and floor areas as extracted from historical data. In Ghana, Frimpong et al. (2003) and Asiedu et al. (2014, 2015) cited financial constraints, referring to delays in honouring certificates for payment as leading to cost overruns. Therefore, the delivery partners in terms of the source of funding were considered, as a factor to ascertain whether making payment within the stipulated time would affect the completion cost. Five delivery partners were considered: the first project was delivered with the used of foreign donor funds and partners (F). In most cases funds were readily available and payment was done on time. There are also non- governmental delivery partners who access their funding readily and do not necessarily follow the government-stipulated procurement routes regimentally (NGO); the Ghana Educational Trust Fund (GETfund) which obtains funding allocated through parliament and has inconsistencies with payment as a result of unnecessary delays; internally generated funding from governmental agencies or institutions' who undertake their own projects with more flexibility in the release of funding and prompt payment options; and lastly, the government of Ghana sponsored projects (GOG) from the consolidated fund which is known for delays in payment. There are lump sums in contracts that occasion underestimating of the actual cost of items they were provided for; hence the assumption is that the prevalence of this will give rise to unexpected expenditure resulting in cost overrun. Furthermore, the tendering route or type from the work of Love et al. (2012) was considered as a risk factor that contributes to overrun of cost and as such was included as a variable in the model.

Liquidated damages (LAD) as part of the contracts terms that pushes for completion of projects on was considered on the basis that it compels contractors to complete projects on time and prevent them from suffering any financial penalties. A mobilization fund, if found in the contract, was assumed to make the contractors' liquidity good and remove to some extent fluctuation on the amount advanced, thereby reducing the uncertainty associated with price escalation (inflation) leading to a reduction in cost overrun. The classification of contractors, which usually is done, taking into account capital for work by the contractor, equipment availability and holdings, knowledge or experience of contractors and personnel employed, was used in the model on the grounds that the type of contractor would influence how they perform.

2.5 Project life cycle or plan of work

Project phases or life cycle refers simply to how works are planned; it seeks to rationalize how best to organize the processes involved in a construction project. Increasingly various construction industry bodies across countries produce and keep reviewing plans of work, mainly to overcome organizational difficulties allied to delivering projects. As the plan of works are varied owing to the fact that they are from various institutions, Huges (2001) poses the question as to which plan, as varied as they are, will be the most suitable for all projects. Furthermore, the construction industry is vastly fragmented. Moreover, the key players along the project phases or life cycle are numerous and they have diverse interdisciplinary professional backgrounds.

There are different stages of the project lifecycle as classified by various institutions across the industry and this research would like to depart from enumerating them but would rather adopt the summary and comparison in the 5^{th} edition of the code of practice for project management for construction and development as depicted below in Table 2.1.

Table 2.1: Comparison of project life cycles by professional bodies.

Code of Practice for Project Management	Royal institute of British Architects (Plan) of Work 2003	BIM Digital Plan of Work 2013	BS 6079-1:2010	ISO 21500-2012
1. Inception	0. Strategic definition	1. Strategy	1. Conception	1. Initiating
2. Feasibility	1. Preparation and brief	2. Brief	2. Feasibility	2. Planning
3. Strategy	2. Concept design	3. Concept 4. Definition	3. Realization	
4.Pre-Construction	4. Developed design 5. Technical design	5. Design		3. Implementing
5. Construction	6. Construction	6. Build & Commission		Controlling
6. Testing and commissioning				
7. Completion, handing over and occupation	7. Handover and close out	Handover and close out	7. Operation	
8. Post-occupation review and in use	In use	8. Operation & end of life	8.Termination	5. Closing

Adopted from CIOB 5th Edition Code of Practice (2014)

2.6 Project life cycle

As can be deduced from the project life cycle or plan of work from these codes in the Table 2.1, the plans vary yet they all seem to point to the entirety of projects. However, to put cost overrun issues into plans of work one must be mindful of the arguments put forward as to what actually constitutes cost overrun. Ahiaga-Dagbui et al. (2014) opined that there should be a change in the perception and measuring of overruns: they argued that it is simple to put the difference between cost at inception and completion as cost overrun. They further stated that depending on the route of procurement of projects, cost can be put in in three phases, namely cost growth, underestimating and overrun. This argument is based on the plan of work.

What is important to this end is whether we get the cost right. Whether it is cost growth or underestimate or overrun, clients, especially in the public sector, will want to know what they are committing to at the beginning. The project delivery cycle supports the arguments above. In reality, some routes of procuring projects make these stages occur concurrently as can be observed from the comparison of the institutions' plans of work as discussed earlier, however more projects follow the plans of work as they have been carried out in distinct stages. If cost is what is important as far as cost overruns are concerned this work contends that arguably there are four costs that can be observed during project delivery. These are cost at conception stage, cost at full definition stage, cost at contract award stage, and cost at completion stage. If this is true, Odeck (2004: 4353) suggests, "It should be demanded that all cost estimates be accompanied with some thorough sensitivity analysis. The idea is to enlighten the policy makers on the possible consequences for risks involved" and this can be achieved by considering a review of cost at the four stages as being put forward by this research. Reviews at the point of the first three of four cost stages could broaden understanding on overrun issues or serve as a guide or a project dashboard in steering the affairs of projects. The factors in cost overruns can only be properly used when they are aligned with the cost determinant factors during estimating. It is therefore important to look at the theoretical concepts on cost and cost estimating.

2.7 Theoretical concept on cost and cost estimating

Cost is usually as a financial worth of energy, quantifiable resources, time and consumables, risk suffered and opportunity sacrificed in the creation and provision of goods and services. By extension this monetary value, even if estimated, is usually required by construction clients any time they want to execute a project. It is of great concern to all clients in the construction industry to produce a facility that is functional, of the required quality, and delivered within acceptable budget and duration. Hence they are all concerned with the cost of projects. This construction cost that is of importance must be well monitored throughout all the stages of the life cycle of the project.

In realising the cost, construction practitioners use cost estimating to establish the likely or estimated cost of projects. This process of estimating tries to make use of both mechanical and subjective expert knowledge in ascertaining the likely total cost of projects, thus providing useful information for decision-making. The precision of estimating construction costs of projects is pivotal to the success of the project. This process is often referred to as cost modelling. Akintoye & Fitzgerald (2000) referred to defined it as "a technical process"; Asthworth (2008) suggests cost modelling is a "technique"; Seeley (1983) looked at it as a "procedure" while Ferry et al. (1999) says it is "a symbolic representation". What is unique and important about these definitions is that they all acknowledge that the process leads to establishment of the estimated cost: whether a technique, procedure, symbolic representation or a technical process, the aim is to arrive at the estimated cost of a project. It is therefore necessary for research of this type to look at the existing cost-estimating models in the body of knowledge in the construction industry.

2.8 Cost-estimating models

The development of theoretical models over time was classified by Ashworth (1999) as follows:

Between 1960 – 1970:

Traditional or single-point deterministic model: This model uses cost per m^2, judgment, approximate quantities and elemental analysis.

From 1970 – 1980:

Mathematical models: Delphi techniques, expert judgment, parametric modelling resources and process modelling are used in this classification.

From 1980 – 1990:

Value-related models: They include risk analysis, the Monte-Carlo simulation and life cycle costing techniques.

From 1990 – 2000:

Integrated knowledge-based models: Price, in-house and other proprietary systems evolved within this period.

From 2000 – current date:

Fortune & Cox 2005 extended the work done on previous models in Skitmore & Marston's seminal work (1999) by adding newly emergent models and classifying them into the following four categories:

2.8.1 Traditional models

These models rely heavily on comparing data on previously completed projects using one or both of the means such as geometrical and spatial arrangement, and similarities and differences in function of projects. They make use of the traditional style of bottom-up estimating method based on the standard method of measurement. Superficial floor area methods, bills of quantities, the cubic method, and functional unit methods are some of the standards methods for the estimating process.

Akintoye & Fitgerald (2000) suggest that the traditional model is the most widely used in the construction industry. The wide usage may be as a result of features such as established standards, repetitive usage, and the non-complex nature of these models. Fortune & Cox (2005) suggested that the traditional models are popular in usage, especially in the UK. Construction professional bodies such as the Royal Institute of Chartered Surveyors (RICS), the Institute of Civil Engineers (CES), and the Canadian Institute of Quantity Surveyors (CIQS), among others, developed and constantly review standard methods of measurement for use in estimating traditionally. These models provide in-depth information on construction activities and classify work or trade activities into standards tables and sections, thereby allowing uniformity of quantifying works that is consistent and unique among practitioners who use their standards in the industry.

Reviews are regularly conducted by these institutions to incorporate current trends and practices of the industry into these standard methods of measurement. These reviews allow for better classification of cost-estimating similarities of functions, information from existing projects, and best practices to inform future decisions. Conducting research to develop an improved model in this study would amount to extending the boundaries of cost-estimating models.

2.8.2 Mathematical model

This model adopts life cycle costing models, parametric models, and econometric or statistical models in estimating the cost of projects. Approaches to estimating the cost of projects are usually based on statistics and regressions analysis (Kim, Ann & Kang, 2004). Kim et al. (2004) describe the regression model as having a well-defined mathematical basis as well as the ability to measure how well a curve matches a given data set.

Skitmore & Patchell (1990) acknowledge that regression is a very powerful statistical tool and is of good use in both the analytical and predictive examination of potential new items that contribute to the overall estimate reliability. Tam & Fan (1999), however, suggest that regression is not very appropriate when it comes to description of non-linear relationships that are multidimensional and consist of multiple input and output problems. However, this model has no specific approach that will help estimators choose the cost model that best fits the historical data (Bode, 2000; Bode, 1998; Adeli & Wu, 1998; Garza & Rouhana, (1995) and Bode (2000) suggest that it is always necessary to review the variables influencing the estimating in advance and it is difficult to use this method when dealing with a large number of input variables.

2.8.3 Knowledge-based model

Ashworth (1999) enumerates the expert system, case-based reasoning and price as part of knowledge-based models. Marzouk & Ahmed (2011) gave an overview of case-based as a system that embodies expertise from past cases. In each case there is a description of the problem and the relevant solution or outcome. The knowledge and reasoning process used by an expert to solve problems is implied in the solution. Problems are matched to previously solved ones and similar ones are retrieved. These methods are able to adapt a retrieved solution when used in a different problem-solving context. Kim et al. (2004) refer to case-based as an alternative to the expert system and as it is based on experience, and is useful for a wide variety of problem-solving tasks, including design, diagnosis and planning as mentioned by Kolodner (1993). This model is used to solve problems by first observing the unique features of a problem, secondly, by identifying these unique features in previously solved similar problems and lastly, by predicting the direction of the new problem on the basis of the similarly experienced problems with adjustments.

2.8.4 New-wave models

The advancement of research in artificial intelligence paved the way for the development of new-wave models. These models include fuzzy logic, algorithms, and artificial neural networks (ANNs). Further developments encompass computer-aided design (CAD) models, among others. Ahiaga-Dagbui & Simth (2012) concluded that developed ANN models have several potential applications in the construction industry. They can be easily converted to desktop packages for use by practising professionals to predict the final cost of projects. The use of ANNs could reduce the time and resources spent in estimating and serve as well as a benchmark against detailed estimates. It also provides alternative solutions. ANN models are a viable option for forecasting costs of construction as they eliminate the establishment of a good cost-estimating relationship that mathematically describes the cost of a system as a function of the variables that have the most effect on the cost of that system. Alex, Hussein, Bouferguene & Fernando (2010) suggest that the use of ANNs is on the increase for several reasons, namely the ability of ANNs to model interdependencies, to deal with non-linear relationships that co-exist between cost-related parameters and they can handle incomplete data sets more effectively. These are critical in achieving realistic cost.

What is of more importance is the introduction of a newer and more advanced algorithm called extreme learning machine (ELM). ELM has been proposed recently for training single hidden layer feed forward neural networks (SLFNs) and it aims at overcoming the drawbacks and difficulties faced by the traditional algorithms (Huang, Bai, Z., Kasun, & Vong, 2015). It has the advantage of using less training time, tuning, and has an improved speed (Huang et al., 2012). Compared to the others ELM is efficient and tends to reach a global optimum (Huang et al., 2015). It has the advantage of efficiency and can be generalized on a wide range of problems across different fields of discipline (Zhang, Song, Wang, & Hou,2014; Huang Zhu & Siew 2006). It tends to reach the smallest error, the smallest norm of weights, has good generalisation performance and runs extremely fast. In the quest of improving the estimating accuracy and developing a good prediction model, this research intends to use ELM as opposed to the traditional multiple linear regression to determine whether there is an improvement in the comparison of the models and to see how improved the ELM can be

Why Extreme Learning Machine was chosen for this modelling comparison based on the following key considerations:

High number and type of variables

There were at least 12 variables to model during the prototyping stage of this research. Most of these variables had at least 3 options to choose from. Tendering type, for example, had options of competitive, sole source and highly quotation whiles contractor type had options of D1K1, D2K2, D3K3 and D4KA, Furthermore, there was a wide mix of types of variables and their scales of measurement. In particular, most of the variables were categorical, instead of the more usual continuous type. Huang (2015) suggests that Extreme leaning machine cope better with categorical variables, the curse of dimensionality and multicollinearity, statistical conditions where two or more variables are highly correlated or dependent on each thereby resulting in spurious predictions when both of those variables are included in the model (Hair *et al.,* 1998).

Non-linear relationships and training time.

Extreme leaning is a sophisticated modelling technique that is capable of modelling extremely complex functions. Linear modelling, typically regression, has been the commonly used technique in most modelling domains since linear models have well-known optimization strategies. These linear functions usually take the form of y= mx + c. However, where the linear approximation is not valid (which is frequently the case) the models suffer accordingly. Instead of trying to model all data to a "best-fit" line in regression equations, extreme learning machine attempts to model the non-linear relationship between the variables, in a better manner. Traditionally, all the parameters of feed-forwarding neural networks need to tune and need a very long time to learn. However, the extreme learning algorithm provides smaller training error and performs better. It only needs to turn parameter.

2.9 Conclusion

Cost is among the cardinal concerns of clients. Projects must be delivered within an acceptable cost. Thus estimating techniques must be improved upon to realize this all-important goal.

There is overwhelming evidence of cost overruns in the construction industry. It has cut across continents and goes deep into the fabric of projects in all nations; from the developed to the developing and even to underdeveloped nations who also share in this problem. Numerous research studies attest to this.

Research of all kinds tried to determine the root causes of cost overruns in the construction industry and these causes have been classified in order to find solutions. Classification in terms of project life cycle vis-a-vis the causes through the lens of the major cost variance at the stages of the project life has not been considered. A critical look at this very area coupled with a review at this stage would go a long way in providing new insight and could possible lead to an improved way of handling construction cost overruns.

This chapter has presented an overview of the level of cost overrun issues in the construction industry. It shows that cost overruns cut across all continents, and are prevalent throughout all nations, be they developed, developing and under-developed nations. The level of project complexity is not different to cost overrun, whether big or small. The root causes of cost overrun were identified to include risk, strategic misrepresentation, and optimism bias. Further causes beyond these are technical inconsistencies. Cost overruns run through the life cycle of the projects as well. Theoretically, cost estimating evolves from the traditional model, mathematical model, knowledge-based and now new-wave models. What is of concern is that cost overruns are still prevalent in the industry. To address this, new methods have to be employed continually to improve estimating. Hence it is important to use extreme learning machine, which has not been used previously in construction cost prediction research, since it needs less training time compared to other popular new-wave methods such as regression and artificial neural networks.

Chapter 3: Cost overrun in the construction industry – Ghana, Nigeria and South Africa

3.1 Introduction

This chapter describes the construction industry as it exists in Ghana, Nigeria and South Africa. Specific attention is given to public sector project executing agencies. The structures of the construction industry, legislation governing public procurement and tendering procedures and managing trends are discussed. Highlights of the challenges facing the sector are also covered and the overrun and its related issues pertaining to South Africa and Nigeria are discussed as well.

3.2 Overview of Ghanaian construction industry

The construction industry in Ghana employs about 632,736 workers and has output constantly increasing from GH¢ 1016.30 million in 2006 to GH₵ 3036.96 million in 2016 (GSS, 2016). The industrial statistics section contributed about 30.6 per cent to the country's gross domestic product (GDP) in the 2015 MFEP (2016) budget.

The main characteristics of the housing stock in Ghana today are overcrowding in small rooms, exorbitant rents (approximately 45% of a household income), and the declining quality of residences, among others. The housing deficit requires an average of over 500,000 units annually to provide adequate housing within the next 10 to 20 and is estimated to be 1.7million in the year 2017. (Center for Affordable Housing Finance in Ghana 2016). This gives credit to the sector as a key contributor to the country's development goals. Table 2.1 provides information of the sector's contribution to Ghanaian economy from 2013 to 2016.

Table 3.1: Growth contribution of construction sector to Ghana's industry

Year	Growth contribution from the industry
2013	287million
2014	288million
2015	2950million
2016	2036.926million

Source: Ministry of Finance and Economic Planning, (2016); Ghana Statistical Service (GSS).

3.3 Participants in the Ghanaian construction industry

According to Ayeribi (2005), the government of Ghana is the major client of the architectural, engineering and construction industry as in many countries. It regulates the industry using four main ministries, namely the Ministry of Roads and Highways (MRH), the Ministry of Water Resources, Works and Housing (MWRWH), the Ministry of Trade and Industry (MTI) and the Ministry of Ports, Harbour and Railways (MPHR).

Government agencies such as the Ghana Highway Authority (GHA), the Department of Urban Roads (DUR), the Public Works Department (PWD), the Department of Feeder Roads (DFR), the State Housing Corporation (SHC), the Architectural and Engineering Service Limited (AESL), the Building and Road Research Industry (BRRI) and Getfund have been set up to support the ministries to perform their duties.

The Local Government Act 462 of 1993 further gives mandate to Metro/Municipal /District Assemblies (MMDAs) to act as planning authorities in settlement planning schemes in Ghana.

A few prominent accredited professional associations such as the Ghana Institutes of Engineering (GhIE), Ghana Institution of Surveyors (GhIS), the Ghana Institute of Architect (GhIA) and private consultancy firms the construction industry together with private contractors also occupy strategic positions in the industry.

3.3.1 Ministries

The Ministries have the specific task of guiding and coordinating executing agencies such as the Ghana Education Trust Fund, the Department of Urban Roads and the Department of Feeder Roads in addressing the transition from large bureaucratic agencies undertaking road works basically by force to leaner, commercially oriented organisations, outsourcing virtually all road projects to the private sector. It also addresses decentralisation issues (particularly pertaining to DFR and to some extent the DUR) with a gradual shifting of planning and operations to lower regional or local levels in compliance with overall government of Ghana policies. The following ministries and agencies carry out works that are related to the building sector.

3.3.2 The Ministry of Water Resources, Works and Housing (MWRWH)

The **MWRWH's** functions include the formulation and co-ordination of policies and programmes for the systematic development of the country's infrastructure requirements in respect of works, housing, water supply, sanitation, and hydrology. The Ministry co-ordinates and supervises related agencies and private sector participants to realise its goals.

To carry out its mandate, the following institutions and organisations have been created under the Ministry for the implementation of its policies and programmes:

3.3.3 Architectural and Engineering Services Limited (AESL)

The AESL was originally set up to provide consultancy services for all public works. It was known then as the Architectural and Engineering Services Corporation (AESC). Under the government of Ghana's divestiture programme, the AESC was privatised in 1995, and its name changed to AESL. Its roles still include the provision of consultancy services for government projects: however, like any other private firm, it now tenders for consultancy services in order to cater for its existence.

3.3.4 Public Works Department (PWD)

The Public Works Department was set up during the colonial days to provide a wide range of services including roads, buildings, water and electricity supply. The Department is currently mainly responsible for the programming and co-ordination of the rehabilitation, maintenance and reconstruction of public buildings and estates. It serves the DAs and only provides consultancy services for MDAs' smaller jobs.

There is also a Technical Services Centre under the MWRWH that operates as a Resource Centre and Project Management Unit. It is responsible for the administration of World Bank-funded programmes for the improvement of urban development. In other words, it offers assistance to other Government Ministries in their dealings with the World Bank and other donors' development projects such as roads, drainage, water supply, electricity and sanitation.

3.4 Source of funding

The public and private sectors in general fund construction projects. In Ghana, the funding for the ministries comes from the following sources:

Consolidated Fund: This is revenue generated by the government through taxes and levies and allocated to the various sectors of the economy to run their projects.

Donor Fund: This incorporates loans, credits, and grants received from bilateral and multi-lateral donor agencies. It usually comes through agreements with the government. These funds are made available for specific projects. The donor agencies include the World Bank, International Monetary Fund (IMF), European Union (EU), African Development Bank (AfDB), Japanese International Co-operation Agency (JICA) and others.

Road Fund: This is generated by collecting a levy on fuel or petroleum products meant for vehicle usage, fees from vehicle licensing and road use, and fees from international transit. These funds are used solely to support road projects (Amoah et al., 2011).

3.4.1 Ghana Education Trust Fund (GETfund)

In August 2001 parliament enacted an Act to provide finance to supplement the provision of education at all levels. The Act mandates a value added tax (VAT) service to deduct two-and-half per cent (2½%) from the prevailing VAT tax and pay it directly into the GETfund. The fund is meant to support public institutions under the Education Ministry to develop and maintain their academic facilities and infrastructure.

3.5 Procurement system and contractor selection in Ghana

The type of procurement system used in any project delivery will emanate from the client's point of view or be based on the advice given to the client by his/her professional advisors (Ashworth, 2008). However, Ghana has been and still is using the traditional system of procurement to a large extent (where the design is separated from execution) with a few new methods of procurement.

The execution of projects mostly follows different stages of the project cycle but in some cases the design and execution stages can be done in a single procurement operation. Thus they clients, request the services of consultants who produce designs together with all necessary contract documents and later engage the service of a contractor to undertake the construction itself. It is worth mentioning that whereas the public sector is restricted in the choice of procurement system or method and the way it can select a contractor, the private sector is not.

Despite there is no directive in the procurement act insisting on lowest bid in Ghana, there remains the tendency of government agencies and department using the lowest bidder as a means of selecting contractors to execute their contracts in order to satisfy the general public.

Ministries, departments and agencies (MMDAs) were using different procurement documents based on convenience until the Public Procurement Act 663 (PPA) of December 2003 was passed.

Owing to the existence of numerous donor funds in Ghana, there are numerous conditions of contract that include FIDIC conditions of contract, conditions of contract for building and civil works for the Ministry of Works and Housing, the World Bank's standard bidding document for the procurement of works and other procurement guidelines of various donor agencies.

3.5.1 The Public Procurement Act (Act 663)

This Act was passed to address the inefficiencies in public procurement over the years. It sets out the legal, institutional and regulatory framework within which all public projects must be procured. It further covers the procurement of goods, works or services and the disposal of stores and equipment.

The Public Procurement Act (PPA) establishes the five basic pillars of public procurement, namely a comprehensive, transparent, legal and institutional framework, clear and standardised procurement procedures and standard tender documents, an independent control system, procurement staff proficiency, and anti-corruption measures (World Bank, 2003).

The PPA comes in nine (9) sections, namely the establishment of the board, procurement structures, procurement rules, methods of procurement, tendering procedures (subdivided into three parts), methods and procedures to engage the services of consultants, review, disposal of stores, plant and equipment, and miscellaneous provisions. Whereas the Act applies to procurements that are financed wholly or partly by public funds, international donor agencies such as the World Bank and others still use their preferred set of procurement procedures in administering projects that they sponsor.

According to Harris and McCaffer (2006), the aim of any procurement is to achieve, among others, quality of the built environment, sustainable and sensitive land use and best value for money in all other aspects. Therefore, all procurement decisions are expected to be based on vigorous assessment of all the available options in each set of circumstances throughout the life of a contract through effective contract monitoring and control. This is the policy of the government of Ghana as well considering the limited resources available to her as a developing country.

The best value for money is reflected in the key principles of the PPA, namely economy and efficiency, competition, transparency and accountability. Its main objective is to establish and maintain high-level efficiency in public procurement process. The PPA entreats all stakeholders to ensure that public sector procurement is carried out in an open, fair, consistent, efficient, and competitive manner. The Act explicitly requires public procurement activities to aim at the following:

- Ensuring that the government's requirements for goods, works, services, and disposal of stores and equipment are met through open and fair process that provides a high degree of competition and value to the economy of Ghana;
- Ensuring that all bidders have reasonable notice and opportunity to bid;
- Fostering national economic development by giving every capable Ghanaian the opportunity to do business with the government;
- Encouraging Ghanaian businesses to be competitive and sustain quality product development;
- Adhering to international agreements which are not at variance with national laws covering relations with Economic Countries of West African States (ECOWAS), development partners and other countries that create economic opportunities for Ghanaians; and
- Being accountable to the public for procurement decisions.

On the section of method of procurement, the Act stresses competitive tendering while giving room for two-stage tendering, restricted tendering, single-source procurement and requests for quotation. However, there are strict procedures for how the variants could be used.

Tendering procedures, including invitation of tenders and application to pre-qualify, submission of tenders, evaluation and comparison of tenders, are fully covered. Methods and procedures to engage the services of consultants that may include the possibility of a project manager are also stipulated in the procurement for easy use.

Furthermore, the Act specifies the number of people on each possible tender board and states the threshold for each procurement method. It further stipulates the threshold of contract that can be awarded for each approval board for the ministries, MMDAs, and departments.

3.6 Managing Trends of Construction Projects in Ghana

According to Smith (2002), unlike the developed countries, many developing countries do not have a mature construction industry (the trend of contract management is not well defined). In Ghana, direct measures undertaken by government to ensure smooth contract management include the issuance and monitoring of expenditure ceilings for each MMDA consistent with budget and updated cash flow forecasts. Also, all procuring entities must seek clearance from the Ministry of Finance, through certification of proof of the availability and adequacy of funding, before any works contract is awarded.

The ministries are the nominated employer in contract relations and the executing agencies that fall under them are responsible for contract management. Per the PPA the tender boards do the award of contracts after evaluation and recommendations by the executing agencies. Sometimes, depending on the nature and extent of works, external consultants are hired to do the evaluation and recommendations. However, the variation orders that arise must be granted by the executing agencies.

For example, whiles the Ministry of Roads and Highways is the employer, there is a contract division within the GHA responsible for contract management. They have a site operation manual where procedure for the approval of interim payment request is also detailed. Monthly and quarterly reports are issued on work in progress in a standardised format in the site operation manual as well.

In the case of DFR, the contract documents are finalised and approved by head office while regional units handle the procurement procedures. However, works funded by foreign donors have specific guidelines. The bid evaluation is done by their regional units and must be validated at the head office. Progress reports and interim payment certificates are prepared together by the various regions and sent to the head office, where payment requests are made by the sector ministry.

The DUR have a Project Management Manual that details their procurement process and is intended for use by MMDAs. Their payment system is similar to that of the DFR.

In the educational sector the GETfund makes allocations (must be approved by parliament) to educational institutions yearly. These institutions have development committees that decide on the types of projects that are needed. The projects are managed by the consultancy firms (selected through competitive tendering) that provide the design and the construction itself is carried out by contractors.

Regulation of construction activities by the Government is done through sector ministries. The public sector has been the frontrunner in determining the rules and conditions that govern the execution of projects.

In the area of capacity building from time to time these agencies run short courses and specialised training for contractors on procurement related issues as well as construction management. In consultancy firms either in government set-ups or in the private sector, the project manager position tends to rotate among the disciplines as mentioned before and the project must be a member of a related discipline's professional institution in Ghana.

3.7 Challenges facing Ghanaian construction industry

In Ghana, the challenges that exist in the construction industry are not very different from those the construction industry is facing globally (Addo-Abedi, 1999). However, Smith (2002) attests to the fact that the developing countries for which

Ghana is not an exception have their challenges compounded by socio-economic stress, persistent resource shortages and institutional and legal weaknesses.

The culture and other related practices of the environment in which the industry is operating are littered with problems. Some of these problems include government policy and restrictions imposed on public construction projects, bureaucratic procedures, contractual clauses incorporating factors of unfair risk sharing, and inadequacies in control mechanisms. Other problems can be attributed to inadequate technological know-how, lack of research and development, and lack of education and training in new technologies and management approaches (Smith, 2002).

According to Ayerebi (2005), long-term strategic planning by both public and private sectors is difficult and so is the monitoring and control of procurement. Delays in project implementation and execution, and huge contractual payment arrears to contractors and suppliers also exist.

The Ghanaian construction industry is largely fragmented. There are numerous companies that often employ few people, forcing the players involved to depend on survival instead of investing in the future *(*Eyiah*,* 2004*)*. Development of human resource, inadequate capacity building and lack of competitiveness of the local firms are other challenges facing the industry. As Ayerebi (2005) pointed out, there exists a substantial perception gap between the required standard by consultants and employers and the capacity of the contractors or their site agents to match. This is compounded by the international contractors' and firms' unwillingness to develop their competitors (local firms).

Failure to adopt best practices working in other countries and lack of interface between the industry and the research institutions to implement the large database of knowledge on the Ghanaian construction industry are equally problematic. Different conditions of contract exist. The MMDAs maintain separate lists for the pre-qualification of contractors and use different standard conditions of contracts for works procurement. World Bank-administered projects, on the other hand, use

FIDIC conditions of contract for works and shortlists for the selection of consultants.

According to Eyiah & Cook (2003), the classification by the Ministry of Works and Housing is too general and obsolete and the registration criteria, lists of contractors and monetary thresholds are not updated regularly. Most government-financed architectural consultancy services and project supervision have been assigned to the Architectural and Engineering Services Limited on a sole basis (World Bank, 1996). Another challenge is a situation where a single contractor buys and prices all the bidding documents. Contracts and suppliers register under different company names and end up being awarded a number of contracts (Crown Agents, 1998; Westring, 1997).

Players in the industry often encounter difficulties in processing claims and, for fear of not getting a job in the future, withdraw these claims. Consultants, contractor and agencies delivering works and services to government establishments try to make a profit by cutting corners or abandoning the work. There is poor contract management and training and the working conditions of the construction workforce are inadequate (World Bank, 2003).

The lack of modern project management skills and practices hinder the effective performance of the industry, leading to improper risk management.

Finally, suspicion of foul play is prevalent among the citizens. This is pronounced by the perceived polarization of the sector (Eyiah, 2004*)*. Agencies are seen to be awarding contracts based on political party affiliation, ignoring competence and the resource base of the contractor.

The construction industry in Ghana, just as in any other developing country, contributes meaningfully to its national economy. It furthers employs many workers at a rate that is ever increasing. However, there are still housing deficits in the country.

According to Ayerebi (2005), the Ghanaian government is the major client in this industry. Thus, it controls and regulates the activities of the industry. This is done by using the infrastructure-related Ministries that have supporting agencies implementing the policies and plans of the government. There are a few private sector participants together with professional bodies of the various construction-related discipline undertaking projects as clients in the Ghanaian construction industry but majority of the construction projects are funded by public at the detriment of the private sector. Funding of government projects comes from the Consolidated Fund, Donor Fund, Road Fund and GETfund.

The traditional system of procurement is widely used in Ghana and the passing of the Public Procurement Act 663 was meant to address the inefficiencies in the system. Therefore, most of the activities regarding procurement have been regulated by the Act. However, there are variations to the procurement routes and options as far as some projects are concerned. Most often these are donor-funded projects.

The management of projects has largely been influenced by the Act and is still traditional. This is because there is one main construction-related university augmented by ten regional polytechnics that produce the human resources for the industry. These institutions have no or limited output of human resources in terms of project management.

According to Addo-Abedi (1999), the challenges that exist in the construction industry in Ghana are not very different from those the construction industry is facing globally. The sector is faced with, among others, the following problems: government policies and restrictions imposed on public construction projects, bureaucratic procedures, contractual clauses incorporating factors of unfair risk sharing, and inadequacies in control mechanisms. Players in the industry often encounter difficulties in processing claims and, for fear of not getting job in the future, withdraw these claims. There is a lack of modern project management skills and practices, thus hindering effective performance of the industry, leading to improper risk management.

3.8 Cost overrun and related issues

3.8.1 Causes of cost overrun

To some extent factors of cost overrun were investigated across the construction industry in Ghana. The findings suggest similar patterns in most of the works done in the public sector. The reason for the concentration on public sector projects might be due to an earlier notion that Ghana, being a developing nation, has the government as the major client in the construction industry. Notable among the research conducted on factors of cost overrun are the following:

Asiedu & Alfen (2015) investigated the "extent and leading events" that caused cost overrun. They were of the view that the factors are within the system (endogenous) and prominent among them are the planning of projects that are poorly done, payments delayed beyond the contractually accepted time, non-enforcement of conditions of contract, and political meddling. Unnecessary price increase payments (for labour and material) and changes decided by clients were the factors also identified in the work of Frimpong et al. (2003) and Furgar & Agyarkwa-Baa (2010). Constraints in terms of finances and lack of proper planning were found to be the fundamental causes of cost overrun (Asiedu & Alfen, (2014).

Among the important factors of cost overruns, the first five is persistent non-payment of periodic amounts due to contractors, non-payment of the prevailing interest rate (locally) on delayed payments by government, prolonged processes in payment that lead to delay, inability in securing funds beyond the budgeted amount, meddling, and procurement process littered with practices uncalled for (bribery and fraudulent selection of contractors).

3.9 South African experience – Cost overruns in South Africa

3.9.1 **Introduction**

Change is emerging in the African construction industry. Projects are becoming more complex and bigger. Deloitte's African Construction Trends report (2013) indicates that the contribution of the industry to economies cannot be down played.

According to the capital expenditure by public sector institutions in South Africa for 2016, provincial statistics on new construction works from 2015 to 2106 are shown in Figure 3.1 below

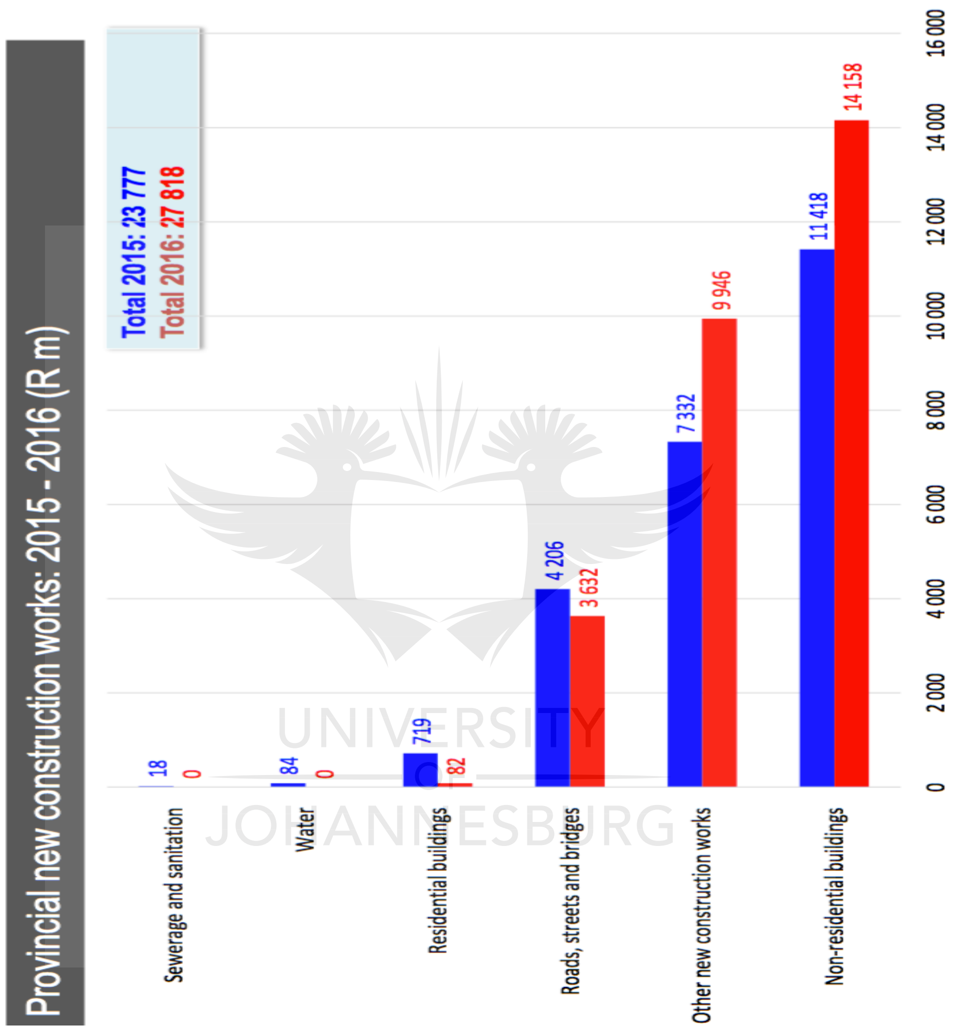


Figure **3.1:Provisional statisics on new construction projects in the public sector of South Africa**

Adopted from Deloitte's African Construction Trends report (2013)

However, in that same report, there is an account on extra budgeting capital expenditure for 2016. This confirms that the existence of overrun factors still lingers in the industry in South Africa.

The construction industry according to a government report (cidb CIIs,2015) comprises residential buildings, roads, streets and bridges, other new construction works and non-residential buildings.

There is the emergence of infrastructural development in the deprived communities as a government intervention. This trend is solving in one way or the other the unemployment situation and boasting economic activity. The Public Works White Paper (1997) suggests the sector engages over 400 000 people, and helps in addressing the unemployment situation.

The advent of democracy brought about mixed economic performance, accordingly Baloyi & Bekker (2010) stated that the best performance of the economy occurred between 1999 to 2003. Thus the country enjoyed a growth rate that has been on par with global average trends. However, this economic performance has not yet yielded the needed employment growth to match the increase in the populace and unemployed.

As the construction industry contributes immensely to the economy of the country, the South African government regularly utilizes it to stimulate growth and aid the economy to recover from recessions. Large capital, often taxpayers' money, is used on construction projects and yet they overrun to large magnitudes thus, Baloyi & Bekker (2010) called close monitoring of performance of public projects in terms of time and cost. The contribution of this industry cuts across various fields and in some cases companies such as PricewaterhouseCoopers often try to capture the monetary benefits enjoyed by various stakeholders. The construction industry development board (cidb)'s construction industry indicators. (CIIs) 2015 summary report stressed the need for improvement in the sector and refers to the level of client's satisfaction as

The level of satisfaction of a client with a contractor's performance on a project is an important indicator (or measure) of the contractor's ability to execute and complete a project within the required expectations of the client. "Feedback is the food of champions", and it is important for contractors to get feedback from clients on their projects so that they can improve their performance on future projects.

These CIIs represent an aggregated industry view of the satisfaction of clients, contractors and the client's agent.

Projects overrun everywhere; the South African construction industry had specifically shared in overrun and concern for early delivery. The needed infrastructure for the 2010 FIFA World Cup was being delivered together with the global recession that followed right after. Winning the bid for the 2010 FIFA World Cup event gave rise to concern as to whether the country would be able to deliver the infrastructure needed for the event within budget, despite its ability to complete large projects over the years (Baloyi & Bekker, 2010).

Several construction projects are delayed beyond expectation and consequently exceed preliminary cost and time estimates. Style of management, non-reviews of project designs, inadequate planning, lack of motivating workers, delayed payment, and the quality of management during the project phase are among the factors that impact project delivery negatively as observed by Olatunii (2010) and these factors cut across the municipalities of the provinces in South Africa.

3.9.2 Public sector infrastructure trend

Infrastructure (2017) reports the trends in public infrastructure spending from the 1998/99 to 2015/16 budget year: more than R2.5 trillion was spent on infrastructure in the public sector. The expenditure went up from R48 billion in 1998/99 to R261 billion in 2015/16, an average annual increase of 6.8 per cent after discounting inflation. Mostly companies state-owned companies have been the biggest sponsors in public sector expenditure as per the duration concerned, contributing R1.1 trillion in total. Municipalities and provincial departments have also increased infrastructure spending, contributing R500 billion and R580 billion respectively for the construction of schools, hospitals, clinics and other community-related infrastructure.

3.9.3 Construction projects overrun in South Africa

Infrastructure overruns are predominant in Africa as well: nearly R30bn is being spent on two of the government's infrastructure projects (Transnet and Eskom) where questions are being raised about its ability to undertake new projects and the level of oversight over the organizations managing the projects (BdLive, 2013).

Important projects in South Africa have experienced significant cost overruns as was suggested by Ramabodu &Vester (2013). The Soccer City Stadium (Johannesburg), Green Point Stadium (Cape Town), and Moses Mabhida Stadium in Durban experienced cost overruns of 174%, 483% and 267% respectively. Global Labour Column (2013) suggested the budget for the stadium construction and related infrastructure was pegged at R2.3 billion. Surprisingly though as at 2010 the estimated total likely cost was R39.3 billion, an immense 1709 per cent escalation as per the original estimate. The cost of constructing a stadium increased from the original estimate of R1.5 billion to the newest estimate of over R17.4 billion, amounting to a 1008 per cent increase.

The Medupe project, according to the estimate of professionals in the industry, was expected to exceed R130bn (including interest) and to exceed the deadline by 48 months. Infrastructurene.ws (2014) attested that the cost overrun would be R77.1 billion.

The first stage of Transnet's new multi-product pipeline (NMPP) from Durban to Johannesburg experienced an increase in cost of R23.4bn when the initial cost was R9.5bn at the start of the project (over a phase of six years).

Eskom's Ingula Pumped Storage Scheme project in the Drakensberg overran the cost to R23.8bn from an initial estimate of R8.9bn (in six years) according to BDLive, (2013).

3.9.4 Studies on causes of cost overrun in South Africa

The Global Labour Column (2011) suggested that the numerous cost overruns issues are the result of bid rigging. According to them, five major companies were the main constructors both on the FIFA World Cup Stadia construction and related infrastructure projects. These companies were investigated in 2007 by the Competition Commission of South Africa on the basis of collusion and uncompetitive practices regarding projects delivery. In a further study by Baloyi & Bekker (2011), they classified factors causing overrun as external factor-related, client-related factors and contractor-related factors. However, on the FIFA World Cup (2010) projects fluctuation in prices and escalation in material cost were the highest ranked factors. Inaccuracy in estimating, lack of skilled labour, and delay in award of contracts, among others, were also identified as factors causing overrun. Ramabodu & Vester (2013) suggested scope changes, inadequate design during tendering, claims, variation order and mismanagement issues as the main causes of cost overruns.

The works of Olapado (2007) attested that in specifications and scope variations, originated predominantly by clients of projects and to some extent by consultants on project are the most widespread bases of cost escalations. This variation comes from changes in clients' financial capability, modifications in owners' specifications, errors in design and inadequate time for contract documents' preparation.

Ramabodu & Vester (2013) advocate the necessity of identifying factors that influence construction costs and duration overruns at the inception of the project so as to ensure a reduction in cost overruns, thereby improving the performance of cost.

The key contributors (CIDB, 2011) to the poor quality of construction projects resulting in cost overruns in South Africa are possibly to be procurement-related hurdles. These hurdles include underhand dealings and exploitation, 'political interference', the route used by clients, the desire for low bids, the preference for particular firms, and inadequate information to enable the selection of consultants and contractors based on the right qualification criteria.

Research conducted by Mukuka, Aigbavboa & Thwala (2014) on the perceptions of construction experts on the causes and effects of cost overruns in Johannesburg concluded the following as the principal ten causes of cost overruns in Johannesburg: contractors' inexperience on projects, poor project management, lack of adequate planning, contractors' incompetence, insufficient financial provision, shortage of skilled site workers, poor workmanship, inaccuracy in estimates, complexity of projects, and site conflicts. Further investigation into the causes of delay and cost overruns in engineering procurement, construction and management projects in South Africa conducted by Nkobane (2012) reveals thirty-nine potential causes of cost overruns, namely increase in material prices (as a result of economic and political conditions), construction delays, inconsistency in supply of material and equipment, monopoly by some suppliers, currency instability, erratic weather conditions, untimely decision making by government, breakdown in coordination, flawed or inappropriate choice of site, labour instability, non-existence of organizational activity to reduce materials waste at site, wrong scheduling of workers on site, design modifications, added work as directed by clients', late changes during construction, amendments to specified drawings at construction stage, inadequate bills of quantities, wrong quantity take-off, poor review of standard in drawings and contract documents, lack of team work by professionals at the design phase, unfinished design while executing tender, amendment to drawings while constructing, inexperience of local regulators, interruption in construction, suppliers' delays, lack of commitment of donors, bad decision by clients, tendency of donors' favouring lowest bid, resources limitation, unavailability of funds, underhand dealing during tendering, insufficent cost forecasting and supervision through stages of projects, claims associated with extension of time, insufficient cost reporting, variation order delays, changes in government policies on projects, and delay in taking over the site.

From their work in Northern Cape public projects of South Africa, Gaetsewe, Monyane & Emuze (2015) presented the major causes of cost overruns as ambiguous scope of work, inexplicit design brief, lack of co-ordination due to inadequate time, work scope changes due to poor planning during the design stages, information on design coming in late, lack of quality workmanship resulting in rework, contractual claims accompanying extension of time, material fluctuation, incompetent selection of contractors and consultants by project owners, unfamiliarity of similar works by contractors and consultants, lack of capability in terms of human resource and finances, and corruption exercises during the procurement process. The causes of cost overruns in South Africa are similar to those of Ghana.

3.10 Construction industry of Nigeria: Cost and cost overrun matters

According to Kasimu (2012), on-going cost overrun in the construction industry is one of the key issues Nigeria as one of the developing countries in West Africa is facing. The construction industry remains an important contributor to the nation's economy: while the contribution is less than the manufacturing or other service industries, it cannot be ignored (Aibinu & Jagboro, 2002). In south-east Nigeria, the industry has gone through intricate changes in contemporary times to the extent that industry players nowadays are looking to embrace numerous endurance tactics to accommodate the seemingly intense competition so as to execute and finish projects within or on the stipulated time and cost (Ubani, Okorocha & Emeribe, 2013)

The economy of Nigeria has been progressing by an average of 7 per cent a year, as against an average of 3 per cent in South Africa. According to Cameroon Tribute (2009), this may come from its prevailing robust investment scheme, with the advantage of its large population, considerable natural reserves and strategic location. The output of the construction industry to the growth of nations demands enhanced productivity in the industry in terms of cost and time management that would result in savings for the country economy (Aibinu & Jagboro, 2002). Corporate Nigeria (21/2011) attests to the fact that Nigeria has the world's largest construction hub. Despite the fact that the global economy is recession, the industry in Nigeria is growing rapidly and will possibly grow even faster over the next ten years.

Lamodi (2015) suggested that the economy of Nigeria was acknowledged as the biggest economy in Africa as of the year 2014. Among the leading drivers of economy, construction industry is regarded as the one that drives the country's development. This stem from the fact that practically all other segments of the economy somehow depends largely on its products and services to perform their core functions. According to Dantata (2007), since Nigeria's economy is one of the largest in Africa, there is the chance for it to be among the world's strongest with its abundance of natural and human resources in the country. Unfortunately, the Nigerian economy continues to over-rely on the oil sector.

This construction sector is the second significant industry that engages human resources, after that of food production. It is the key source of employment and employs about 70 per cent of the labour force in Nigeria. Thus the construction industry's contribution has monetary effects as it deals with a large capital flow, together with fairly large labour resources, according to Amusan (2011). There is therefore the need to management these resources as far as the delivery of projects is concerned as quality management of these resources largely determines the success of projects.

Approximately 2.5 million labourers participated directly in construction projects during the 1990s. Of these the breakdown is as follows: 88% were unskilled or had limited levels of skill, 11% had within medium to high levels of skills while the rest, namely 1%, worked at managerial level. The variance in the percentages of skills poses a serious skill requirement problem in Nigeria. Thus the construction industry has issues with skills shortage. Work in the sector tends to be more physical and labour intensive, and poorly paid, thereby discouraging young people from being willing to work in the sector (Oyewobi, Ibironke, Ganiyu & Ola-Awo, 2011).

3.10.1 Cost overruns in Nigeria

Ameh, Soyingbe, & Odusami (2010) point out that accounts of the construction industry globally abound with examples of cost and time overrun of completed and uncompleted projects. In 1994 a study investigated 8,000 projects and the results suggest only 16% of the projects were completed within the confines of the procurement triangle, namely cost, time and quality.

Ameh et al. (2010) emphasize that all of the projects currently undertaken in developing countries such as Indonesia, Nigeria, Saudi Arabia and Malaysia experience major construction overruns in terms of cost and time, according to studies undertaken.

Construction projects in South-Eastern Nigeria, according to Ubani et al, (2013), experienced abandonment and impediments as result of the Nigerian civil war. Kaming et al. (1997) indicate many studies on cost and time overruns have been carried out. The findings show that cost overruns outweigh time overruns. Thus Kasimu (2012) suggests is a worthwhile to look into the problem of cost overruns.

Cost, time and quality, and lack of performance have been innate features of construction projects in Niger State of Nigeria, according to Mamman & Omozokpia (2014).

Oyewobi et al. (2011) affirmed that overruns as far a cost and time in projects are concerned have become common phenomena in the Nigerian construction industry in modern times. Project sponsors, executors and owners are not getting value for money. However, changes have been taking place in all sectors of the economy over the last ten years and construction is no exception.

On evaluating the impact of factors on project cost in terms of overruns in the construction industry of Nigeria, Da & Jagboro (2007) revealed that risk factors have the highest impact on projects completed; these account for 30% of the cost overrun of the contract sum projects undertaken using the project management mode of procurement. Procurement through the traditional method, which is commonly used, for procurement in Nigeria indicated 16% of cost overrun of the project contract sum related to risk factors; design and build 9%; management contracting 25% and labour only 13%. The rest of the impact factors found are physical, environmental, design, logistics, financial, legal, political, construction, operation and time schedule slippages. Financial and political factors were found to be the most significant in construction projects cost overruns

3.10.2 Causes of cost overruns in construction projects in Nigeria

Construction projects in Nigeria experienced on-going cost overrun and the issue requires urgent attention in the industry (Kasimu, 2012). It more prevalent where traditionally contracts are awarded to the lowest bidder, especially with projects undertaken in the public sector. In developing countries, including Nigeria, the trend is to use the lowest bid strategy (Ubani, Okorocha & Emeribe, 2013).

Amusan (2011) maintains the key factors that cause cost overrun in construction projects are insufficient planning, lack of experience of contractors on projects, economic issues such as inflation, persistent variations, design changes, complex issues associated with projects, fraud and unrealistic contract periods (too short).

According to Ameh et al. (2010), the findings of a study conducted in Nigeria indicate inexperienced contractors, high material cost, price changes in material, recurrent changes of design, an unstable economic situation, excessive interest rates on loans and unfavourable repayment terms, corrupt practices and bribes were the leading factors causing cost overruns.

The findings of Mansfield et al. (1994) established that the key factors causing overrun in project are financial issues, lack of proper contract management, material unavailability, price changes, and underestimating. The market situation, lack of work experience, time, material price changes and political interference were found to be significant factors causing overruns in the industry in Nigeria, according to Kasimu (2012).

From the work of Ubani & Ononuju (2013), factors causing overrun are changes in government, corruption practices, terms of payment, political interference, leaking engineers estimate, improper project planning, insufficient feasibility studies on project viability, incomplete design, lack of technical know-how, and inadequate risk analysis.

According to Amusan (2011), the probability of projects experiencing cost overrun is 90%, and this can result in the abandonment of projects. The next highest cause is the owners' capital being tied up. Furthermore, there is the likelihood of cost overrun due to payments and unconstructive time, while the possibility of contractors going bankrupt or insolvent or not making a profit stands at 70%.

Changes in policies and governments from a ruling party to another create delays in terms of payment for works done. Changes in scope, prices inconsistencies of working materials and equipment, unrealistic design and error in design, inconclusive technical feasibility, amendment and reworks are factors suggested by Ubani & Ononuju (2013). They further pointed out that if these are ignored, it leads to abandonment of projects. Aibunu & Jagboro (2002) state claims (either as a result of delay by clients or fluctuation) lead to significant cost overrun in projects in Nigeria. Furthermore, incomplete information on design or excessive change in design hinder the smooth progress of projects, thus leading to time overrun, disputes, litigation and arbitration, and finally, to abandonment of projects.

3.11 Conclusion

This chapter dealt with the cost and cost overrun issues in the Ghanaian construction industry, as well as in the construction industries of South Africa and Nigeria. It was seen that in Ghana the government is the major client in the industry. Private client participation is evolving. It was further pointed out that there are structured ministries and agencies that undertake business on behalf of the government as client. It was also seen that the funding of government projects is from both domestic and foreign. There is a structured procurement system that governs government business. Cost overrun cuts across almost all government agencies and through most projects. There are delays and abandonments, among others. The issues of the cause of cost overrun in Ghana are not very different South Africa and Nigeria; more so, estimating practices still need improvement in Ghana. The chapter further examined South African and Nigerian cost overrun dynamics it obvious that Ghana needs improvement as far construction cost prediction is concerned. There are little to no cost prediction exercises and this area must be pursued vigorously.

Chapter 4: Methodology

4.1 Introduction

Modelling the cost of construction projects is mainly quantitative in terms of research methodology. This assertion is because the modelling process involves great usage of numbers. The data, mostly numeric, used in the process were the result of activities of human resources: it is the people behind these that make the difference in any output that the data presents. Simulations using data are only as good as the operators and the environment that generates it. Thus it is a matter of necessity to take a closer look at the context and participants that operate in it so as to better understand interpreting outcomes from simulations. It is therefore necessary to use qualitative means (open-ended questions) to seek the view of the participants in order to study their environment and elicit information directly. This will enable the researcher to derive useful meanings from evidence gathered and to deduce issues from data correctly.

This aspect of this research presents the objective behind the usage of a mixed method and explains the theories guiding the modelling using multiple regression and extreme learning machine.

4.2 Approach to research

The decision to use this type of research approach would have to take into consideration the kind of problem to be investigated and the possibility of effectively dealing with the questions of the investigation. Quantitative and qualitative are the predominant research approaches (Fellows & Liu, 2008), thus the quantitative research approach associates with positivism and aims to find data that is factual in nature in order to examine trends between facts and how their trends relate with theories and previous findings in research literature. It is essentially a scientific investigation in that theory and relevant literature-specific issues drive the aims and objectives, enabling suggestions and postulations to be put to the test by probing the correlation between variables that differ from each other.

To use this approach Creswell (2009) suggests considering the following three characteristics of the problem under study: the elements that influence an outcome must be recognized, the possibility of measuring an intervention, and having full knowledge of the preeminent predictors of a result. Essentially, to use quantitative approach the problem should be measured, the causation must exist, as well as the ability to generalize and reproduce same application in other situations (Bryman, 2012).

The measurement has to do with the possible computation of the measure of correlation between perceptions or factors. Causation is the ability to establish the trend between a group of factors and their effects. Deductions and finding from a sample should be extended to large samples (generalization) and reproducing means the possibility to use the same method, for other different subject areas. For the work to be credible, reproduction is important because results from other investigation must be consistent with earlier deduction and findings.

Studies that use qualitative approach, on the other hand, typically do not have prior postulations: their preoccupation is to advance theories through the study of participants by interacting with them. The qualitative approach seeks to investigate participants' views on different problems or the attributes participants assign to social occurrences or people's cultures (Creswell, 2009).

The qualitative approach is ethnographic in that the researcher examines an integral social faction in its natural environment over a lengthy duration by gathering data from observation or interviews. It implies that strategies such as discussions would be analysed, either written or recorded, to unearth meaning and knowledge to formulate theories. The work of Fellows & Liu (2008) suggests a thorough investigation (that is qualitative) usually serves as preface for studies that are quantitative.

Beyond these two approaches to research, there is pragmatism which claims knowledge can come out of people's actions, events or conditions and results rather than from antecedent conditions. There is focus more on the problem rather than the methods and an investigator should use all approaches necessary to better understand the research problem at hand. Thus, this system is not committed to any one system of philosophy or reality. Consequently, the mixed approach to research supports this ideology in that researchers draw freely from both quantitative and qualitative postulations when conducting research. The significance of this mixed method is the combination of both quantitative and qualitative features to sufficiently tackle a specific research problem. It tends to allow observation of the problem instead of the art of gathering data and analyzing them in the case of the two traditional approaches (Cresswell, 2009). This allows the investigation to be more robust in its approach rather than using one of the two traditional approaches (qualitative or quantitative).

This study's focus is predicting the cost of construction projects by developing predicting models, thereby using a principally quantitative approach. However, it adopts the usage of qualitative methods as well, chiefly at the preliminary stages of the work. Relevant literature on construction cost overrun issues has been explored and reviewed, Further works were carried out by conducting an explorative interview with a professional who has a great deal to do with projects from inception to completion (from whom outfit data were retrieved). The interview enabled the researcher to focus on the specific context of the Ghanaian construction industry in order to understand and collate information personally about the historical and cultural settings of the players and the completed projects from which data was extracted within the industry. This step laid a firm foundation for variable selection and executing the modeling process to the point of developing a probable construction cost- estimating model.

4.2.1 Research framework

The structure used in the research has been presented in in the first chapter. The work began with an exposition of the central purpose of the research aims, objectives and questions and continued with a review of relevant literature on construction project cost overruns to align the issues with the project life cycle. This laid the foundation for investigating the issues in the industry in order to know the participants and the specific context in which they operate. This threw more light on the issues and paved the way for efficient data collection. This led to subsequent analyses, model development and validation.

There are not many predicting cost estimating models relating to building works in the construction industry. However, research, especially in the civil works of the construction industry, relating to cost or duration prediction abounds. These works investigated the factors that contributed to the estimation of construction cost and overruns in the construction industry. These estimating models used mainly regression, case-based reasoning and artificial neural networks. Regression was used on road projects to analyze project cost overruns, delays in time and variation orders. In the work of Bordat et al. (2004) seven factors were used; these include the amount of bid, project location, competitive bids, initial duration, the weather, consultant estimates and variance in competitive bids which were used in their model. Artificial neural network and regression were used to predict variances in cost of projects that are reworks in their nature. (Attalla et al., 2003). In all, 36 factors were seen to have contributed to cost performance. It was realized that artificial neural network performs better than the regression model and responds better to larger numbers. A regression model was developed for overruns in the highway construction sector (Odeck, 2004). Apart from variation between the estimated and actual cost seen, it came to light that overruns are more common in smaller projects as compared to larger projects. Neural network has mainly been applied in construction cost estimating (Ahiaga-Dagbui, 2014; Boussabaine, 1997; Moselhi et al., 1991).

4.3 Qualitative research

The study begins by identifying the factors that affect construction cost overrun in Ghana. Most of the factors that contributed to cost overrun were chiefly identified from existing literature. These factors, which were identified from literature, needed to be validated before being used in the construction industry in Ghana. This was mainly done using a qualitative strategy, namely an explorative interview. Qualitative research seeks to explain rather than measure. An un- structured interview was used to identify the factors affecting cost overrun as well as linking them to the factors identified in literature. The interview was more focused on explaining what causes construction cost overrun as identified in literature.

4.3.1 Research methodology

The results of the interview and documentary findings from seven (7) consulting firms consulting, twelve (12) construction firms and four (4) tertiary institution development directorates in Ghana are reported. In-depth interviews with the quantity surveyors, architects, contractors and project managers who do the actual estimating and pricing in contracting firms in Ghana were carried out in 2017.

Each session of the unstructured discussions was documented with the interviewee's permission and lasted an estimated 75 minutes. The consultants were from registered firms in Ghana with their personnel registered with the various professional bodies such as the Architectural Council of Ghana, the Ghana Institution of Surveyors and the contractors were mainly those registered with the Ministry of Works, Water Resources and Housing. These firms, tertiary institutions and contractors were chosen based on their set up, namely a well-resourced office, good professional background of their staff, the size of their organisation and projects they have carried out. The above considerations made it possible to assume that they have had enough experience with projects and their delivery. All the consulting firms, construction firms and tertiary institution directorates have been operating in Ghana for at least the past 18 years. The projects carried out in these entities were mainly the construction of new building works that are education related, with some refurbishments. The quantity surveyors, architects, project managers and contractors have an average of 15 years' experience. Their outfit employed an average of about 540 staff who work either directly in the offices or at the project locations. They include managers, engineers, architects, quantity surveyors, administrative staff and skilled and unskilled labourers. The study used convenience sampling to select the participants for the interview across each consulting firm, construction firm and the tertiary institution. However, those selected were part of the estimation process of construction projects executed in their various organisations. In all, 69 participants who were mainly quantity surveyors, architects and project managers, were interviewed.

The review of relevant literature reveals the fundamental areas and predictors to look out for. The principal reason for using an unstructured interview approach was to understand the context within which the projects were delivered and to allow the respondents to mention factors they thought attributed to the cost overruns in the construction industry. The primary description of factors emerged from the analysis of the responses this guided the retrieval of useful factors from completed projects for the development of models. The interpretation was much by the way they referred to the factors. The interview was analysed by looking out for common points in description and aligning those points with available data extracted from completed project cases. Contract documents, including interim and final certificates, variations orders, final and interim reports, site-meeting minutes and tender evaluation reports, were collated and examined. In some cases, designated officers were available to explain and give further details. The facts obtained through the interview were checked with collated contract documents.

The variables obtained from explorative interview, discussions with seasoned practising personnel from the construction industry in Ghana, together with a review of relevant literature made it possible to identify and contextualize twelve variables that contribute to the unexpected cost for the completion of projects. These are contract sum at award stage (Asum), initial duration (ItD), consultant estimate (CEsum), number of floors (NFl), and gross floor area (GFa), which were extracted as numeric variables. Categorical variables identified in Ghana included type of tendering with three levels (sole sourcing, quotation and competitive bids); funding agents as funder with four levels (internally generated fund [IGF], Ministry of Finance [GOG], Education Trust Fund [GETfund], non-govermental organisations [NGOs]; and foreign donors (F)); region with 10 levels; scope change with two levels (Yes or No); provisional sum; mobilization fund; liquidated and ascertained damages (LAD); and contractor type. Literature suggests that, the scope of projects has a significant contribution to the resultant cost of projects and the time it takes to complete them (Ng et al., 2001; Bromilow, 1969; Walker, 1995). Scope of projects was referred to as the measurement of size in relation to function.

Attributes such floor areas, storey height or number of storeys and building volumes were influential in the scope of projects (Love et al., 2012, 2005). Thus, the more complex and large a project is, the more the possibility of an increase in cost overrun as a result of errors in estimating and variation orders coming from the initial contract sum at the award stage, changes in the number of storeys, and floor areas as extracted from historical data. In Ghana, Frimpong et al. (2003) and Asiedu et al. (2014, 2015) cited financial constraints, that is delays in honouring certificates for payment, as leading to cost overruns, hence the delivery partners in terms of source of funding were considered as a factor to ascertain whether making payment within the stipulated time would affect the completion cost. Five delivery partners were considered, namely first project delivered with the use of foreign donor funds and partners (F) which in most cases were readily available and payment was done on time; non- governmental NGO) delivery partners who obtain their funding readily and do not necessarily strictly follow the government stipulated procurement routes; the Ghana Educational Trust Fund (GETfund) whose funding is allocated through parliament and has inconsistencies with payment as a result of unnecessary delays; internally generated funds from institutions which undertake their own projects with more flexibility in the release of funds and prompt payment options; and lastly the government of Ghana sponsored projects (GOG) from the consolidated fund which is known for delays in payment. There are lump sums in contracts that occasionally underestimate the actual cost of items they were provided for; hence the assumption is that the prevalence will give rise to unexpected expenditure, resulting in cost overrun. Furthermore, the tendering route or type (Love et al., 2015) was considered as a risk factor that contributes to overrun of cost and as such was included as a variable in the model.

Liquidated damages (LAD) as part of the contract as a factor was considered on the basis that it compels contractors to complete projects on time without suffering any financial penalties. The mobilization fund, if found in the contract, was assumed to make the contractors' liquidity good and remove to some extent the fluctuation on the amount advanced, thereby reducing the uncertainty associated with price escalation (inflation) leading to reduction in cost overrun. The classification of contractors which usually is done taking into account capital for work by the contractor, equipment availability and holdings, knowledge or experience of contractors and personnel employed, was used in the model on the basis that the type of contractor would influence how they perform and therefore contribute to unexpected expenditure on projects.

It was found out that there are three costs: consultant estimate, the contract sum at award, and the final cost at completion. Ahiaga-Dagbui (2014) suggests that the change in cost estimate at the two main project phases (preliminary phase and the project definition phase) are better regarded as underestimation while the changes in the construction stage as the real cost overruns that occur. In reality, there are factors such as limited or lack of information, optimism bias and strategic misrepresentation as concurred by Flyvbjerg (2002), and change in scope as project definition progresses and risk as identified by Odeck (2004). What is difficult is that consulting historical data would not reveal most of these factors and not all consultants would be willing to accept some of the factors as contributing to cost overrun in a specific project. This research thus looks at predicting completion cost, taking into account two kinds of overrun as follows: Change in final completion cost and contract sum at award as percentage and change in final completion cost and consultant estimate expressed as a percentage. The assumption is that at any point where these two estimates are arrived at predicting the overrun will produce the difference in cost that must be added to the estimate arrived at to predict the final cost at completion or the actual sum required to complete a project.

4.3.2 Variables for modelling

Research done so far suggests there is no standard baseline for determining construction cost overruns which is vital in predicting cost as far as modelling is concerned, thus leading to variations in the magnitude of overruns and inaccurate estimating models. There are models that take changes in the magnitude of the estimated budget at the preliminary stage of projects and the actual construction cost at completion while others use the difference between the initial contract sum at award stage and the actual construction cost at completion (Xiao et al., 2002). Ahiaga-Dagbui (2012) contended that cost overrun reference point could be misleading and would lead to overvalued and inflated cost variance since the estimate at the preliminary stages may not fully take into account the scope and the character of the project at award stage. Actual construction cost was defined by All-Momani (1996) as the total of all cost at final completion. Hence, the cost overrun was referred to as the change in the actual cost and the budgeted cost expressed as a percentage (Cantarelli et al., 2012). These make the base line to which cost overruns are referred not specific.

According to consulting practitioners and data from the industry in Ghana, there are at least three main emerging costs: the budget, the consultant estimate, the contract sum at award stage, and the final cost of completion. They surface at the various stages of the project life cycle. What is of importance is that there is an element of design to cost where budgets are set as caps for consultants to design to cost or fit the budget ceiling. This act reduces the cost of estimates to three, being consultant estimate, the contract sum at award and the final cost at completion. Ahiaga-Dagbui (2014) suggests that the change in cost estimate at the two main project phases (preliminary phase and the project definition phase) are better seen as underestimation while the changes in the construction stage as the real cost overruns that occur. In reality there are factors such as limited or lack of information, optimism bias and strategic misrepresentation as concurred by Flyvbjerg (2002), change in scope as project definition progresses and risk as Odeck (2004) identified. What is difficult is that consulting historical data would not reveal most of these factors and no consultant will be willing to accept some of the factors as contributing to cost overrun in a specific project. This research thus looks at predicting completion cost taking into account two kinds of overrun as follows: Change in final completion cost and contract sum at award as percentage, and change in final completion cost and consultant estimate expressed as a percentage. The assumption is that whether from the point of consultant's estimate or the contract award sum, there will be a difference in cost at the final completion. This difference must be predicted and added to the initial cost in terms of consultant's estimate or contract award sum to arrive at the final completion cost.

In identifying the predictive variables for the modelling, experience, circumspection and expertise guided the decision. Explorative interviews, discussions with seasoned practicing personals from the construction industry in Ghana together with the review of relevant literature make it possible to identify and contextualize twelve variables that contribute to the unexpected cost for completion of projects.

4.4 Quantitative research

The quantitative section of the study was used to determine the significance of the factors identified in the qualitative research that affected construction cost overrun.

4.4.1 Research methodology

This section elaborates on how the research work is conducted. It also covers research design, approach, data sampling design, variables considered, measurement procedures used, data collection method utilized and how data is analyzed.

4.4.2 Target population

A population in a survey research as defined by Davies (2007) and Groves et al. (2009) is the group of subjects from which the researcher wants to make inferences by using sample statistics. The target population of the thesis comprises all public construction works undertaken from 2011 to 2016 across all the ten regions of Ghana. The projects which are considered included those funded by the government of Ghana, GETfund, IGF, NGOs and foreign donors across the whole country. The population excludes all private construction works carried out in Ghana during the same period.

4.4.3 Sampling technique

Sampling is the process of selecting some elements from a population for analysis. (Mustafa, 2010). The main objective of sampling in quantitative studies is to create a sample which is representative and presents features or characteristics of the population from which they were drawn (Neuman, 2011). The sample that is collected is usually small but representative of the larger population. The results of the sample are therefore used to make an estimation of the entire population. According to Mustafa (2010), the sample size should be small enough to avoid unnecessary expenses and large enough to avoid intolerable sampling errors. Selecting samples is less expensive, less time-consuming, less cumber- some and more practical to analyse.

Sampling techniques are of two types, namely non-probability sampling and probability sampling (Berenson, Levine & Krehbiel, 2009). In non-probability sampling, the samples are selected without any knowledge of the probabilities of their selection while in probability sampling, the subject are selected based on known probabilities. In this research, the probability sampling approach was used because it allows unbiased inferences to be made about the population. Simple random sampling, stratified sampling, systematic sampling and cluster sampling are the four major types of probability sampling. A combination of stratified and simple random sampling was used in this research. The target population was divided into mutually exclusive, homogeneous segments (strata) and then a sample random sampling technique was applied in each stratum.

Advantages of simple random sampling include the fact that each selection is independent of the other selections; each subject in the population has the same and independent opportunity to be included in the study; it is an easier method compared to other methods and also easier to comprehend and communicate to others; and it tends to yield representative samples.

Advantages of stratified random sampling include making it easier and possible to draw inferences within each strata and also do a comparison across stratum; it has slightly lower random sampling errors; and a sample is obtained which is representative of the population because it ensures that subjects are selected from each stratum. In addition, data collection costs may be reduced if the population is divided into homogeneous geographical areas, which facilitates data collections. It allows for analysis to be done within each stratum and makes it possible for different research methods to be applied in each of the different strata.

4.4.4 Sampling and sample size

The main reasons for selecting a sample from population are:

1. Selecting a sample is less costly compared to selecting every subject in the population;
2. Selecting a sample is less time-consuming than selecting every subject of the population; and
3. It is less cumbersome and more practical to analyze a sample than an analysis of the entire population. According to Mustafa (2010), the sample size selected has a direct impact on the accuracy of the results, time and cost of the study. Based on the total population of construction projects undertaken from 2011 and 2016 in each region, a total of 911 samples were selected to enable statistical inference to be made for the population. The samples were proportionally based on the total number per each region.

4.4.5 Data collection

Ghana is mainly divided into ten (10) regions and in order to have a fair representation of data, data on completed projects were gathered across the ten regions. In all, 987 data sets were compiled for use in the model development. In order to develop an efficient and effective model, data gathered needs to be homogenous; thus building projects completed in the education sector of the industry were selected. These projects were executed from the period of 2011 to the end of November 2016.

The processes from the preliminary stages of these projects to the completion stage were consistent with the prevailing procurement act in the country. It was necessary to limit the period of completed projects so as to eliminate the barriers to extracting good data since the industry is known as having poor data housing issues. Contract completion reports, final accounts records, interim and final certificates from archives together with some explanations from consulting personnel, contractors and clients within the construction industry aided retrieval of data for this research.

4.4.6 Data analysis and interpretation

Data analysis and interpretation were conducted using both descriptive and inferential statistics. The descriptive statistics were used to describe the distribution of the data using mean, minimum, maximum and standard deviations for the numerical variables while frequencies and graphs were used to report categorical variables. Inferential statistics were used to make judgments based on the samples collected from the target population. Two different modelling techniques were used in the study: multiple linear regression and Extreme learning machine.

Sequentially the practitioners of the industry were interviewed o ascertain the project delivery process of the historical projects. This laid the foundation for examination of documented projects from archives. Thus the predictive variables were retrieved from these documents for the purposes of modelling. The predictive variables that are quantitative in nature as described in section 4.3.2 were used in the model development.

4.5 Linear regression

Scientists, engineers, quantity surveyors and other professions often collect data in order to determine the nature of relationships between quantities. For example, a quantity surveyor may want to study relationship between construction cost overruns and project duration, location of the project and other factors. It becomes very important to also measure the strength of the relationship between cost overruns and the factors that affect it, to estimate accurately the effect of each factor and to be able to predict accurately cost overrun given a factor. It is also important to determine the change in construction cost overruns for any change in any of the factors that affect cost overruns.

This section deals with the very common problem of fitting a straight line to data. It describes simple or multiple linear regression analysis as a tool for predicting a quantitative response variable. Multiple linear regression analyses are a set of statistical techniques used to assess the relationship between one quantitative dependent (response) variable and several independent (predictor) variables

(Tabachnick & Fidell, 1996).

The objective of a regression model is to predict a single dependent variable from the knowledge of one or more predictor variables. When a single predictor variable is used to predict the response variable, it is referred to as simple linear regression while in situations where two or more predictors are used they are termed multiple linear regressions. Simple linear regression analysis is a statistical technique used to analyse the relationship between a single dependent variable and one independent variable (Hair et al., 2010). In practice, there is often more than one predictor variable. Instead of fitting a separate simple linear regression model for each predictor, a better approach is to extend the simple linear regression model to directly accommodate multiple predictors. Thus a multiple linear regression model uses at least two or more variables to predict the response variable. A simple linear regression model is represented as:

$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i = 1,.....n \tag{4.1}$$

β_0 is the intercept, β_1 is the slope, X_i is the predictor variable and ε_i represents the error term of the random response variable y_i. The X_i is assumed to be fixed and the ε_i are independent random variables with $E(\varepsilon_i) = 0$ and $\text{var}(\varepsilon_i) = \sigma^2$

The multiple linear regression model is given by

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k X_k + \varepsilon_i \tag{4.2}$$

where *p* represents the number of predictor variables. The parameters $\beta_0, \beta_1, ...\beta_p$ need to be estimated.

4.5.1 Parameter estimation

Three general methods of estimation are mostly used. These are the ordinary least squares (OLS) maximum, likelihood (ML) and method of moments (MM) and its extension, the generalised method of moments (GMM). The amount of change in the response variable due to the predictor variables is represented by the terms β_1, ..., β_p which are referred to as regression coefficients/parameters (Hair et al., 2010). The values of the parameters (β_1,...,β_p) are estimated using the least square method (Hair et al, 2010), so that the sum of the squared errors of prediction is minimised. The difference between the actual and predicted values of the response variable, termed the residual (ε), is the prediction error (Hair et al., 2010). Applying the least square method, the slope and the intercept of the model are chosen to minimize

$$S(\beta_0,\beta_1)=\sum_{i=1}^{n}(y_i-\beta_0-\beta_1X_1)^2 \tag{4.3}$$

The parameters β_0 and $\beta_1$are chosen to minimize the sum of the squared vertical deviations or prediction errors.

The equation of the line is given as

$$y=\hat{\beta}_0+\hat{\beta}_1X_1+\hat{\beta}_2X_2 \tag{4.4}$$

The quantities $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$ are called the least square coefficients. The least squared line is the line that best fits the data. The quantity $\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1X_1+\hat{\beta}_2X_2$ is called the fitted value and the quantity ε_i is called the residual associated with the point (x_i,y_i). The residual is the difference between the value y_i observed in the data and the fitted value $y_{i,}$ predicted by the least square line (Navidi, 2006). Points which are above the least square line have positive residuals, and points below the least square have negative residuals. The closer the residuals to 0, the closer the fitted values are to the observations and the better the line fits the data. To compute the equation of the least square line, the values for the slope β_1, and the intercept β_0 that minimize the sum of the squared residuals must be computed. To achieve this, the ε_i is expressed in terms of $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$\varepsilon_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \quad (4.5)$$

The β_0 and β_1 are chosen to minimize

$$S = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 . \quad (4.6)$$

The least square coefficients are derived by taking partial derivatives of S in Equation 4.6 with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$ and setting them to 0. That is, $\hat{\beta}_0$ and $\hat{\beta}_1$ are the quantities that solve the simultaneous equations:

$$\frac{\partial S}{\partial \hat{\beta}_0} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (4.7)$$

$$\frac{\partial S}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n}x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (4.7)$$

The minimisers $\hat{\beta}_0$ and $\hat{\beta}_1$ satisfy

$$n\hat{\beta}_0 + \hat{\beta}_1\sum_{i=1}^{n}x_i = \sum_{i=1}^{n}y_i \quad \text{and} \quad (4.8)$$

$$\hat{\beta}_0\sum_{i=1}^{n}x_i + \hat{\beta}_1\sum_{i=1}^{n}x_i^2 = \sum_{i=1}^{n}x_i y_i \quad (4.9)$$

Solving equations 4.8 for $\hat{\beta}_0$ gives

$$\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n}y_i - \frac{1}{n}\hat{\beta}_1\sum_{i=1}^{n}x_i \quad (4.10)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

Substituting $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$ into equation 4.9 gives

$$\left(\sum_{i=1}^{n} x_i\right)\left(\bar{y} - \hat{\beta}_1\bar{x}\right) + \left(\sum_{i=1}^{n} x_i^2\right)\hat{\beta}_1 = \sum_{i=1}^{n} x_i y_i \qquad (4.11)$$

Solving equation 4.11 for $\hat{\beta}_1$ gives

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}\left(x_i y_i - n\bar{x}\bar{y}\right)}{\sum_{i=1}^{n}\left(x_i^2 - n\bar{x}^2\right)} \qquad (4.12)$$

In simple linear regression model, the estimates $\beta_0, \beta_1, ...\beta_p$ are estimated by the least square method as in simple linear regression. The equation

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + ... + \hat{\beta}_p X_p \qquad (4.13)$$

is called the least square equation of the fitted regression equation. The coefficients $\beta_0, \beta_1, ...\beta_p$ are estimated in order to minimize the sum of the squared residuals. Thus, minimize the sum

$$\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - ..... - \hat{\beta}_p x_p\right)^2 . \qquad (4.14)$$

The minimization is done by taking partial derivatives of Equation 4.14 with respect to β_0, β_1, …….. β_p, and setting them to 0, and solving the resulting p+1 equation in p+1 unknown variables.

4.5.2 Statistical properties of the estimated slope and intercept

The formulas for $\hat{\beta}_0$ and $\hat{\beta}_1$ were derived in the earlier section. From the equations, it is observed that $\hat{\beta}_0$ and $\hat{\beta}_1$ are linear functions of the y_i and therefore a linear function ε_i (Rice 2006). The least-squares estimates possess properties given the assumptions of the linear regression model. These properties are contained in the Gauses-Markow theorem. $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimates of β_0 and β_1. The OLS estimator $\hat{\beta}_1$, is considered as a best linear unbiased estimator (BLUE) of β_1 if the following hold (Gujarah & Porter, 2009):

1. It is a linear function of the response variable in the regression model;
2. It is unbiased, thus, its average or expected value, E($\hat{\beta}_1$) is equal to the true value, β_1;
3. It has minimum variance (efficient); and
4. It is consistent.

Theorem 1:

Under the assumptions of the standard statistical model, the least-squares estimates are unbiased. E($\hat{\beta}_1$) = $\hat{\beta}_1$ for j=0,1,…p, where p is the number of parameters.

Proof

From assumptions E $E(y_i) = \beta_0 + \beta_1 x_i$. From the equation for $\hat{\beta}_0$ In Equation 4.10

$$E\left(\hat{\beta}_0\right) = \frac{\sum_{i=1}^{n} x_i^2 \left(\sum_{i=1}^{n} E(y_i)\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i E(y_i)\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i^2\right)^2}$$

$$E\left(\hat{\beta}_0\right) = \frac{\sum_{i=1}^{n} x_i^2 \left(n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i^2\right)^2}$$

$$E\left(\hat{\beta}_0\right) = \beta_0 \qquad (4.15)$$

Theorem 2:

From the standard statistical model, $\mathrm{var}(y_i)=\sigma^2$ and $\mathrm{cov}(y_i, y_j)=0=$ 0 where $i \neq j$

$$\mathrm{var}\left(\hat{\beta}_0\right)=\frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n} x_i^2-\left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\mathrm{var}\left(\hat{\beta}_1\right)=\frac{n\sigma^2}{n\sum_{i=1}^{n} x_i^2-\left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\mathrm{cov}\left(\hat{\beta}_0,\hat{\beta}_1\right)=\frac{-\sigma^2 \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2-\left(\sum_{i=1}^{n} x_i\right)^2}$$

The variances of the intercept and slope depend on the x_i and on the error variance, σ^2 (Rice, 2009). $\hat{\beta}_1$ is said to be a minimum-variance estimate of β if the variance of $\hat{\beta}_1$ is smaller than or at most equal to the variance $\hat{\beta}_2$ which is any other estimator of β. If $\hat{\beta}_1$ and $\hat{\beta}_2$ are two unbiased estimators of β, and the variance of $\hat{\beta}_1$ is smaller than or at most equal to the variance of $\hat{\beta}_2$ then $\hat{\beta}_1$ is a minimum-variance unbiased, or best unbiased, or efficient estimator.

An estimator $\hat{\beta}$ is said to be a linear estimator of β if it is a linear function of the sample observations. That is, the sample mean is defined as

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n} x_i=\frac{1}{n}(x_1+x_2+......+x_n) \qquad (4.16)$$

is a linear estimator because it is a linear function of the X values. If $\hat{\beta}$ is linear, unbiased and has minimum variance in the class of all linear unbiased estimators of β then it is called a best linear unbiased estimator (BLUE). It often happens that an estimator does not satisfy one or more of the statistical properties in small samples. However, as the sample's size increases, the estimator possesses desirable statistical properties. These properties are referred to as large sample or asymptote properties (Gujarah & Porter, 2009). An estimator $\hat{\theta}$ is said to be an asymptotically unbiased estimator of θ if

$$\lim_{n\to\infty} E\left(\hat{\theta}_n\right) = 0$$

where $\hat{\theta}_{\mathrm{n}}$ means that the estimator is biased on a samples size of n and $\hat{\theta}$ is an asymptotically unbiased estimator θ if its expand value (mean) approaches the true value as the sample size increases. Consider the sample variance of a random variable X.

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

The expected value of S^2 is given as

$$E(S^2) = \sigma^2\left(1 - \frac{1}{n}\right)$$

where σ^2 is the true variance. In a small sample, S^2 is biased but as n increases indefinitely, $E(S^2)$ approaches the true σ^2 . Therefore, it is considered as asymptotically unbiased.

The estimator $\hat{\theta}$ is said to be a consistent estimator if it approaches the true value θ as the samples size increases. The estimator $\hat{\theta}$ is considered to be a consistent estimator of $\hat{\theta}$ if the distribution of $\hat{\theta}$ has zero variance.

4.5.3 Model diagnostics

The multiple linear regression defined in Equation 4.2 can be written much more succinctly using a matrix notation as y = X β + ε. From available data, the estimates of β are found using $\hat{\beta}$= $(X^TX)^{-1}X^T{}_Y$. Improvements in prediction of the response variable are possible by adding predictor variables and in some cases transform to represent aspects of the relationship that are not linear (Hair et al., 2010). In order to achieve this, assumptions are made about the relationship between response and predictor variables that affect the statistical procedure used for multiple regressions. The assumptions are discussed in the next section.

4.5.3.1 Model assumptions

4.5.3.1 Linearity

The linearity of the relationship between response and predictor variables represents the degree to which the change in the response variable is associated with the predictor variable (Hair et al., 2010). The concept of correlation, which is based on a linear relationship, in regression analysis. Linearity of any bivariate relationship is examined through residual plots. Any curvilinear pattern in the residuals will decrease the predictive accuracy of the model and the validity of the estimated coefficients (Hair et al., 2010). When a corrective action is undertaken, the predictive accuracy of the model increases as well as the validity of the estimated coefficients. Linearity is assessed through an analysis of residuals and partial regression plots for each independent variable. The partial regression plots are used to show the relationship of a single predictor variable to the response variable while controlling the effects of all other predictor variables. The use of partial regression plots is more useful where several predictor variables are used in the model because it becomes easy to identify which specific variables violate the assumption of linearity and the needed remedies applied only to them. A non-linear pattern in the partial regression plots for each predictor variable versus the response variable indicates that the assumption of linearity of each predictor variable is met (Hair et al., 2010).

4.5.3.2 Independence of the error terms

Predictions in a multiple linear regression model are not perfect and one will rarely find a situation where they are. The prediction errors should be uncorrelated with each other. That is, each predicted value is independent. Plotting the residuals against any possible sequencing variable can identify the occurrence of autocorrelation. A random pattern in the plots shows that the residuals are independent (Hair et al., 2010).

4.5.3.3 Normality of the error term

The most frequently encountered assumption violation is non-normality of the independent or dependent variables or both (Hair et al., 2010). This is the most fundamental assumption in regression analysis. Whenever the variation from the normal distribution is sufficiently large, all the resulting statistical tests are invalid because normality is required to use the F and t statistics. The severity of non-normality is based on the shape of the offending distribution and the sample size. The shape of the distribution can be described by two measures, namely kurtosis and skewness. Kurtosis refers to the peakedness or flatness of the distribution compared with the normal distribution (Hair et al., 2010). Skewness describes the balance of the distribution. It is also important to consider the effect of sample size on statistical power by reducing sampling error. Larger sample sizes reduce the detrimental effects of non-normality. The simplest diagnostic for the set of predictor variables in the equation is a histogram of residuals, usually with a visual check for a distribution approximating the normal distribution.

Another approach, which is more reliable, is the normal probability plot, which compares the cumulative distribution of actual data values with the cumulative distribution of a normal distribution (Hair et al., 2010). The normal distribution forms a straight diagonal line and the plotted data values are compared with the diagonal. If the distribution is normal, the line representing the actual data distribution closely follows the diagonal. Statistical tests for normality such as the Shapiro-Wilks test and a modification of the Kolmogorov-Smirnov test are used to assess normality. It is important to use both graphical plots and any statistical tests to assess normality. Data transformations provide a means of correcting non-normality and heteroscedasticity. The most common transformations are the logarithm, square root and square. In most instances, all the possible transformations are applied and the most appropriate one is selected.

4.5.3.4 Constant variance of the error term

The presence of non-constant variance of the error term heteroscedasticity is one of the common assumption violations. Statistical tests and residual plots are used for the diagnosis. The presence of a consistent pattern in the plot of residuals versus predicted values indicates the presence of heteroscedasticity. Homoscedasticity is desirable because the variance of the response variable being explained in the dependence relationship should not be concentrated in only a limited range of predictor values. The constant variance of the error term is referred to as homoscedasticity while non-constant variance of the error term is known as heteroscedasticity. This assumption deals with the constancy of the residuals and can be examined using fitted versus residuals plots and the Breucsh-Pagan test.

4.5.3.5 Checking assumptions

There are a number of tools, which are used to check the assumptions of a multiple linear regression.

4.5.3.5.1 Normality

The normality of the error terms in a regression model are usually assessed using a visual examination of the normal probability plots of the residuals (Hair et al., 2010). The quantile-quantile plot is a visual method for assessing the normality of errors. If the probability plot has the appearance of a straight line, the residuals are approximately normally distributed. Other methods such as a histogram and the Shapiro-Wilk test are employed. Another tool for assessing normality assumptions is the histogram of the residuals. The errors are believed to be normally distributed when the histogram appears normal. Histograms and Q-Q plots present a good visual representation of the distribution of the residuals.

However, a formal testing (Shapiro-Wilk) could be implemented. This test provides the value of the test statistics and the associated p-value. The null hypothesis assumes the data were sampled from a normal distribution while the alternative assumes the data were not sampled from a normal distribution. A p-value less than the chosen level of significance (normally 0.05) indicates that the null hypothesis is rejected and there is therefore evidence that the data are not from a normally distributed population.

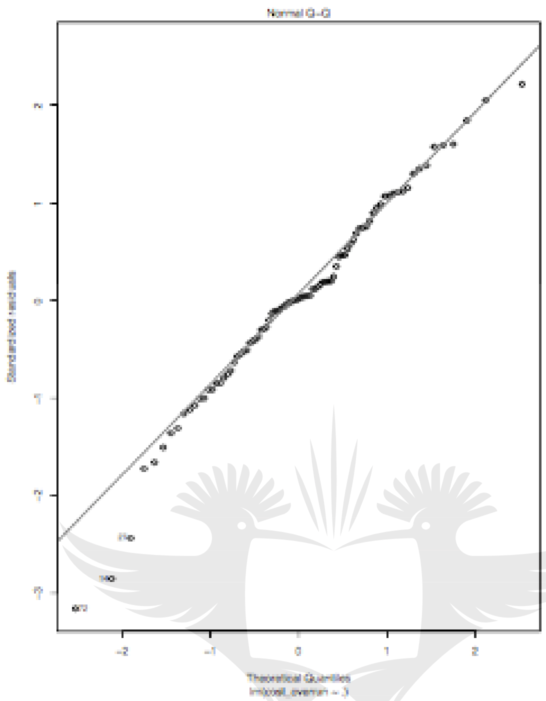


Figure 4.1: The Normal Q-Q plot.

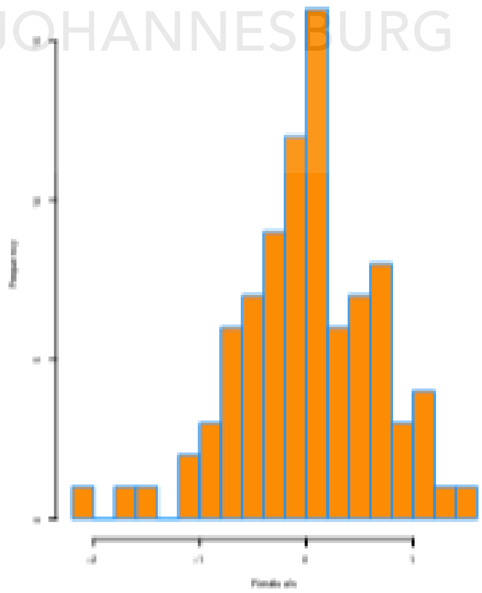


Figure 4:2 Histogram of Residuals

4.5.3.5.2 Homoscedasticity

At any fitted value, the mean of the residuals should be 0 and the variance of the residuals should be constant. If there are no patterns in the plots of the fitted versus residuals plots, it is indicative of homoscedasticity. Even though a fitted versus residuals plot is used to assess homoscedasticity, a more formal test (Breush-Pagan test) is sometimes preferred as well. The null and alternative hypothesis of the Breush-Pagan test is stated as follows:

H_0 = *Homoscedasticity: The errors have constant variance*

H_1 = *Heteroscedasticity: The errors have non – constant variance*

Rejecting the null hypothesis indicates non-constant variance of the error terms.

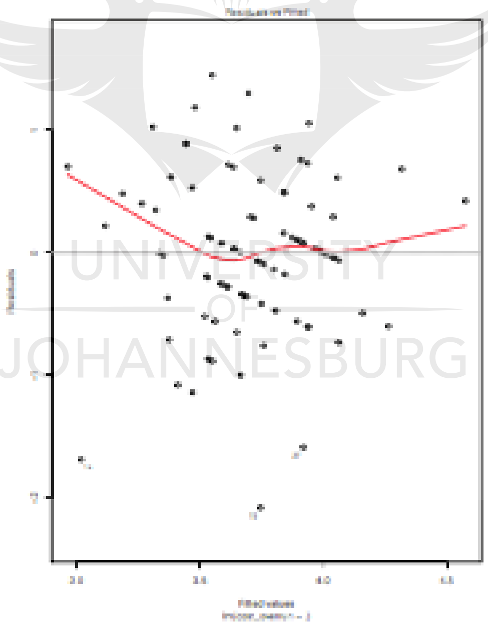


Figure 4.3: Plot of residuals versus fitted values.

4.5.3.5.3 Independence of the error terms

This assumption deals with the effect of carryover from one observation to another, thereby making the residual not independent. The error terms at time t should not be correlated with the error term at time t-1 or any other error term in the past. The

Durbin-Watson test is used to test for independence of the observations. The Durbin-Watson statistic is defined as:

$$d = \frac{\sum_{i=2}^{2}(e_i - e_{i-1})^2}{\sum_{i=2}^{n} e_i^2}$$

The d value always lies between 0 and 4 (Gujarati, 2015). The closer the d value to 0, the greater is the evidence of positive autocorrelation and the closer to 4, the greater is the evidence of negative autocorrelation. If the d value is above 2, there is no evidence of positive or negative autocorrelation.

4.6. Variable selection

Many methods have been proposed for selecting suitable predictor variables in multiple linear regression. Some of these methods include backward elimination, forward selection and stepwise regression (Kuo & Mallick, 1998). These methods sequentially delete or add predictor variables by means of mean squared error. Other methods such as Bayesian information criterion (BIC) and Akaike information criterion (AIC) are used. Variable selection is intended to select the best subset of predictors. This is done to explain the data in the simplest way and remove redundant predictor variables. Unnecessary predictor variables generally add noise to the estimation of other predictor variables and end up wasting degrees of freedom. In some cases, having too many variables results in multicollinearity.

4.6.1 Forward selection

This process starts with no predictor variables in the model and then predictor variables are added one at a time, provided they meet the statistical criteria for entry. The variable that gives the greatest additional improvement to the fit is added to the model (Tibshirani, James, Witten & Hastie, 2013).

4.6.2 Backward selection

Like forward stepwise selection, backward stepwise selection provides an efficient alternative to best subset selection. However, unlike forward stepwise selection, it begins with the full least square model containing all the predictor variables, and then removes iteratively the least useful predictor variable one at a time.

4.6.3. Stepwise selection

This is a combination of backward elimination and forward selection procedure. Variables are added to the model sequentially, similar to the forward selection. However, after adding each new variable, the method may also remove any variables that no longer provide an improvement in the model fit.

4.7 Model selection

There are situations where a large number of predictor variables are used in a model. However, not all the predictor variables in a multiple linear regression necessarily contribute to explaining variation in the response variable. This is referred to as model selection. Also, there is also a choice to be made between models. Good model selection depends on Occam's razor principle, which states that the best scientific model is the simplest model that explains the observed facts (Navidis, 2006). Occam's razor principle implies the principle of parsimony, which states that a model should contain the smallest number of variables necessary to fit the data in terms of linear models. There are three criteria for model selection, namely Akaike information criterion (AIC), Bayesian information criterion (BIC) and adjusted R^2.

4.7.1 Akaike information criterion

This is a measure of the quality of models for a given data. It was first developed by Akaike (1973) as a way to compare different models. For example, it is important to determine what variables influence cost overrun of projects and this may lead to estimating several different regression models. For example, the project duration and location of project may both play a role in determining the cost overruns of a project. Regression models may be run that include duration of project, or any other combination of variables. The AIC is a way to select the best model. The selection of the best model is determined by the AIC value: $AIC = -2\log\{L(\hat{\theta} \mid y)\} + 2K$ where K is the number of parameters, L(θ|y) is the log-likelihood. The best model is the model with the lowest AIC value. The maximised log-likelihood of a regression model can be written as

$$\log L\left(\hat{\beta}, \hat{\sigma}^2\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\frac{RSS}{n}\right) - \frac{n}{2}$$

where $RSS = \sum_{i=1}^{n}(y_i - \hat{y})^2$ and $\hat{\beta}$ and $\hat{\sigma}^2$ were chosen to maximize the likelihood.

AIC is then defined as

$$AIC = -2\log L\left(\hat{\beta}, \hat{\sigma}^2\right) + 2p = n + n\log(2\pi) + n\log\left(\frac{RSS}{n}\right) + 2p$$

The two main components of the AIC are the likelihood (measures goodness of fit) and the penalty (which is a function of the size of the model). The likelihood part of AIC is given by: $-2\log L\left(\hat{\beta}, \hat{\sigma}^2\right) = n + n\log(2\pi) + n\log\left(\frac{RSS}{n}\right)$

The only term that will change is nlog(RSS/n) while then nlog(RSS/n) term is constant across all models. The penalty component of the AIC is given by 2P, where P is the number of parameters in the model. The best model is one with a smaller AIC that has a good balance between fitting well and using a small number of parameters. For comparing models, $AIC = n\log\left(\frac{RSS}{n}\right) + 2p$ is a sufficient expression since $n + nlog(2\pi)$ is constant across all models.

4.7.2 Bayesian information criterion

The Bayesian information criterion (BIC) is similar to the AIC, but has a larger penalty. The BIC quantifies the trade-off between a model, which fits well, and the number of model parameters: however, it generally picks a simpler model than AIC. The BIC was proposed by Schwarz (1978) and was a different penalty. The BIC is defined as

$$BIC = -2\log L\left(\hat{\beta},\hat{\sigma}^2\right) + n\log(n)p = n + n\log(2\pi) + n\log\left(\frac{RSS}{n}\right) + \log(n)p$$

The likelihood of BIC is given by

$$-2\log L\left(\hat{\beta},\hat{\sigma}^2\right) = n + n\log(2\pi) + n\log\left(\frac{RSS}{n}\right)$$

The penalty component of BIC is $\log(n)^P$, where P is the number of parameters in the model. The best model has the smallest BIC value. For comparing models,

$$BIC = n\log\left(\frac{RSS}{n}\right) + \log(n)p$$

is a sufficient expression since $n + n\log(2\pi)$ is constant across models.

4.8 Interpreting the regression coefficients

The estimated regression coefficients represent both the strength of relationship between the predictor variable and response variable as well as the type of relationship (positive or negative). The value of the coefficient indicates the change in the response variable each time the predictor variable changes by one unit while the sign of the coefficient denotes whether the relationship is positive or negative. The regression coefficients play two functions, namely prediction and explanation. Regression involves the use of a model to predict or estimate a single response value. Prediction therefore is an integral part in regression analysis both for estimation and forecasting (Hair et al., 2010). After the OLS method is used to estimate the regression coefficients, the response variable is predicted for the hold out sample. The estimation process sets the weights of the regression coefficients to minimize the residuals (difference between actual value and predicted value). The real benefit of prediction is in forecasting. The regression model is used for prediction using data, which is not used in the estimation process. For example, after validating the model, it is used for forecasting construction costs of other projects. It is important for a regression model to make accurate predictions and also assess the impact of each predictor variable in predicting the response variable. The relative importance or contributions of each predictor variable is examined. Predictor variables with larger regression coefficients would normally make a greater contribution to the predicted response value. However, the regression coefficients do not necessarily provide information on which predictor variable makes greater contributions, as the scales of the predictor variables are often different. It is therefore necessary to ensure that the predictor variables are on a comparable scale. However, regression provides coefficients, which result from the analysis of standardized data. This coefficient eliminates the issue of dealing with different units of measurements and hence reflects the relative importance of each predictor variable.

4.9 Validation of the results

After selecting the best regression model using different techniques, it is important to ensure that it represents the general population and that it is appropriate for the conditions in which it will be used. Empirical approaches to model validation are used, since in most cases prior results are not available. The most common and appropriate empirical validation is to test the selected regression model on a new sample drawn from the general population. However, factors such as cost and time do not permit for the collection of new data. For this reason, the sample is usually divided into two parts: an estimation subsample (training data) for creating the regression model and the hold out or validation subsample (testing data) used to test the model.

The most common standard for comparing regression models is the overall predictive fit (adjusted R^2). The adjusted R^2 is always used because R^2 will increase even if non-significant predictor variables are added. The adjusted coefficient of determination is useful in comparing different regression models involving different number of predictor variables or sample sizes as it makes allowance for the degree of freedom for each model (Hair et al., 2010). Another approach of validating the results is through forecasting with the models. This is done by applying the estimated model to a new set of independent variable values and calculating the response variable. The predicted values of the response variable for the different models are compared with the actual value of the response variable. The model with the least residual (error) is deemed the best model. The confidence intervals of prediction are calculated in addition to the point estimate.

4.10 Conclusion

This chapter gave details of the approach to study. It presents the research framework that has been used. In addition, the methodology adopted was discussed in detail. How data was gathered, the framework for analysis and interpretation were discussed. An overview of multiple linear regression has been given to prepare the way for its usage.

Chapter 5: Extreme Learning Machine

5.1 Introduction

Machine learning and big data analysis have drawn much attention from researchers in different disciplines. Machine learning relies on three key factors, namely powerful computing environments, rich and dynamic data, and efficient learning algorithms (Huang, Bai, Kasun & Vong, 2015). Efficient learning algorithms are highly demanded in big data applications. A number of machine learning techniques such as feed forward neural networks (FNN) have been well used since the introduction of the back-propagation (BP) algorithm (Huang, Huang, Song & You, 2015; Rumelhart, Hinton & Williams, 1986). The traditional BP algorithm is a first-order gradient method for optimisation of the parameters (ie. it involves numerous gradient descent searching steps). It suffers from a slow convergence rate, local minima problem and intensive human intervention (Huang et al., 2015). Various methods have been proposed to improve the efficiency or optimality in training FNN, including second-order optimisation methods (Hagan & Menhaj, 1994; Wilamowski & Yu, 2010), subset selection methods (Chen, Cowan & Grant, 1991; Li, Peng & Irwin, 2005) or global optimization methods (Branks, 1995; Yao, 1993). Although these methods have a faster training speed or better generalisation performance compared to the BP algorithm, these methods still cannot guarantee a global optimal solution.

The extreme learning machine (ELM) has been proposed recently for training single hidden layer feedforward neural networks (SLFNs) and multi- layered feedforward networks and aims to overcome the drawbacks and the limitations faced by the traditional BP algorithm (Huang et al., 2015). The input weight (linking the input layer to the first hidden layer) and hidden layer biases are adjusted in almost all practical learning algorithms of feedforward neural networks. The parameters of the feedforward networks are tuned and therefore there is a dependency between differences layers of parameters (weights and biases) (Huang, Zhu & Siew, 2006).

5.2 Architecture

The gradient-descent method which has mainly been used in various feedforward neural network learning algorithms is generally slow owing to improper learning steps or converge easily to the local minima. Unlike the traditional learning algorithms, e.g. BP, support vector machines (SVM), ELM represents a type of machine learning technique in which the hidden neurons need not be tuned in learning and are randomly generated, thus the hidden nodes could be established before the training data are acquired (Huang et al., 2006; Akusok, Björk, Miche & Lendasse, 2015; Huang et al., 2015; Huang, 2014; Huang, 2015). That is, the hidden nodes in ELM are independent of the training data and independent of each other (Huang, 2014). The hidden node parameters can be generated without the training data. The only free parameters needed to be learned are the connections (weights) between the hidden layer and the output layer. The ELM is thus formulated as a linear-in-the parameter model which boils down to solving a linear system. The ELM, compared to FNN is efficient and tends to reach a global optimum (Huang et al., 2015). The advantages of ELM in efficiency and generalization performance over traditional FNN algorithm have been demonstrated on a wide range of problems from different fields (Huang et al., 2015; Huang, Zhou, Ding & Zhang, 2012; Huang, Zhu & Siew, 2006). The ELM tends to reach the smallest training error, the smallest norm of weights, good generalisation performance and runs extremely fast.

ELM was first proposed for the single-hidden layer feedforward neural networks (SLFNs) (Huang, Zhu & Siew 2006) and then extended to generalized SLFNs where the hidden layer need not be neuron alike (Huang, Zhou, Ding & Zhang, 2012; Huang & Chen, 2007, 2008; Huang, 2014; Huang, Zhou, Ding, & Zhang, 2012; Huang et al., 2015). ELM is an effective and efficient machine learning algorithm which has attracted much attention from different disciplines. ELM is simple in theory owing to the fact that the hidden nodes are randomly generated and the output weight is determined analytically which makes its implementation easy and fast compared to other machine learning algorithms. Even though hidden neurons play critical roles, they need not be adjusted. A suitable number of nodes in the hidden layer is set before training the data and values for the input weights are randomly assigned. The algorithm completes the whole process at once without iterations and generates unique optimal solutions (Mao, Zhang, Liu, Li & Yang ,2014). Neural networks can achieve learning capability without tuning hidden neurons iteratively provided the activation functions of the hidden neurons are nonlinear piecewise continuous functions which include but are not limited to biological neurons.

The most popular learning algorithm used in feedforward neural networks is the back propagation (BP) learning algorithms where gradients are computed by propagation from the output to the input (Huang, Zhu & Siew, 2006). However, the BP algorithm has several weaknesses:

1. When the learning rate is too slow, the learning algorithm converges very slowly. However, when the learning rate is too fast, the algorithm becomes unstable and diverges;

2.The performance of the BP algorithm is impacted by the presence of local minima. The learning algorithm stops at a local minimum if it is located above a global minimum;

3.The neural network may be over-trained by using BP algorithms and obtain a worse generalisation performance; and

4. Gradient-based learning is very time-consuming in most applications.

These problems related with gradient-based algorithms are resolved using an efficient learning algorithm for feed-forward neural networks - ELM. Unlike the traditional function approximation theories which require adjusting input weights and hidden layer biases, input weights and hidden layer biases can be randomly assigned if only the activation function is infinitely differentiable (Huang et al., 2006). The input weights $\mathbf{w_i}$ and the hidden layer biases b_i are not necessarily tuned and the hidden layer output matrix **H** can remain unchanged once random values are assigned to the parameters at the beginning of learning.

5.3 ELM hidden nodes, feature mappings and feature space

ELM transforms the input into hidden layer through ELM feature mappings. The output is generated through ELM which could be either regression or classification (Huang et al., 2015; Huang, Song, Gupta & Wu, 2014; Huang, 2014).

The output function of ELM for generalized SLFN is

$$f(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\beta$$

where $\beta = [\beta_i, ..., \beta_L]^T$ is the vector of the output weight between the hidden layer with L nodes to the output layer with m ≥ 1 nodes; and $h(x) = [h_1(x), ... h_L(x)]$ is the output (row) vector of the hidden layer with respect to the input X. There are many different activation functions which may be used in different hidden neurons. In particular, in real appreciations, $h_i(x)$ can be $h_i(x) = G(a_i, b_i, X), a_i \in R^d, b_i \in R$ where $G(a,b,x)$ is a nonlinear piecewise continuous function which satisfies ELM universal approximation capability theorem (Huang, 2014; Huang and Chen, 2007; Huang, Chen and Siew, 2006; Huang and Chen, 2008) and (a_i, b_i) are the *i*-th hidden node parameters. Some of the commonly used activation functions include:

1. Sigmoid function $G(a,b,x) = \dfrac{1}{1+e^{-(a.x+b)}}$

2. Fourier function $G(a,b,x) = \sin(a.x+b)$

3. Hard limit function

$$G(a,b,x) = \begin{cases} 1, & \text{if } a.x - b \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

4. Gaussian function $G(a,b,x) = e^{-b\|x-a\|^2}$ and

5. Multiquadrics function $G(a,b,x) = \sqrt{(\| x-a \|^2 + b^2)}$

a_i is the input weight vector connecting the input layer to the *i*-th hidden layer, b_i is the bias weight of the *i*-th hidden layer, and β_i is the output weight.

ELM trains an SLFN in two main stages:

(1) Random Feature Mapping and

(2) Linear Parameters Solving.

ELM randomly initializes the hidden layer to map the input data into a feature space in the first stage by nonlinear mapping functions. The random feature mapping of ELM differs from other learning algorithms which use kernel functions or Restricted Boltzmann Machines (RBM) or Auto – encoders/Auto – decoders for feature learning.

Definition 3.1: A neuron (node) is called a random neuron if all its parameters (e.g, (a, b) in its output function $G(a,b,x)$) are randomly generated based on continuous probability distribution.

$h(x)$ actually maps the data from the d-dimensional input space to the L-dimensional hidden layer random feature space (ELM feature space) where the hidden node parameters are randomly generated according to any continuous probability distribution, and thus $h(x)$ is indeed a random feature mapping.

Definition 3.2: A hidden layer output mapping $h(x)$ is said to be an ELM random feature mapping if all its hidden node parameters are randomly generated according to any continuous probability distribution and $h(x)$ has universal capacity, that is

$$\| h(x)\beta - f(x) \| = \lim_{L\to\infty} \left\| \sum_{i=1}^{L} \beta_i h_i(x) - f(x) \right\| = 0$$

holds with probability one with appropriate output weights β

ELM theory aims to reach the smallest training error but also the smallest norm of output weights (Huang, 2014; Huang, Zhu & Siew, 2004; Huang et at., 2005). According to Bartlett's neural network generalization theory (Huang, 2014; Bartlett, 1998; Huang et al., 2012), for feedforward neural networks reaching smaller training error, the smaller the norms of weights are, the better generalization performance the networks tend to have. This may be true to the generalised SLFN where the hidden layer may not be neuron alike (Huang, 2014; Bartlett, 1998; Huang & Chen, 2007, 2008).

Minimise: $\| \beta \|_p^{\sigma_1} + C \| H\beta - T \|_q^{\sigma_2}$ where $\sigma_1 > 0,\ \sigma_2 > 0, p, q = 0, 1/2, 1, 2, ..., \infty, H$ is the hidden layer output matrix

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \cdots & h_L(x_L) \\ \vdots & \ddots & \vdots \\ h_1(x_N) & \cdots & h_L(x_N) \end{bmatrix} \qquad (5.2)$$

C is the parameter controlling the trade-off between these two terms, T is the training data target matrix

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & \ddots & \vdots \\ t_{N1} & \cdots & t_{Nm} \end{bmatrix} \qquad (5.3)$$

where $\|.\|$ denotes the Frobenivs norm.

The ELM training algorithm can be summarized as follows:

1. Randomly assign the hidden node parameters e.g. the input weight a_i and biases b_i for additive hidden node $i = 1, ... L$

2. Calculate the hidden layer output H; and

3. Obtain the output weight vector $\beta = H^{\tau} T$

where H^{τ} is the Moore – Penrose generalized inverse of matrix H, $T = [t_1, ..., t_N]^T$. There are many different methods that can be used to calculate Moore – Penrose generalized inverse of a matrix: orthogonal projection method, orthogonalization method, iterative method, single value decomposition (SVD) (Huang, 2014; Rao & Mitra, 1971; Huang et al., 2015; Huang et al., 2012). The orthogonal projection method can be used efficiently to calculate the Moore – Penrose inverse $H^{\tau} = (H^T H)^{-1} H^T$ if $H^T H$ is non-singular and $H^{\tau} = H^T (HH^T)^{-1}$, if HH^T is non-singular.

It was suggested according to the ridge regression theory (Huang, 2012; Hoerl & Kennard, 1970) that a positive value ($1/\lambda$) can be added to the diagonal of $H^T H$ or HH^T in the calculation of the output weight β. The resultant solution is equivalent to the ELM optimization solution with $\sigma_1 = \sigma_2 = p = q = 2$, which is more stable and tends to have better generalization performance (Tang, Deng, & Huang, 2016; Huang, Zhou, Ding, & Zhang, 2012; Huang et al., 2012; Hoerl & Kennard, 1970).

Thus, in order to improve the stability of ELM,

$$\beta = H^T \left(\frac{1}{\lambda} + HH^T \right)^{-1} T \qquad (5.4)$$

and the corresponding output function of ELM is

$$f(x) = h(x)\beta = h(x)H^T\left(\frac{1}{\lambda} + HH^T\right)^{-1}T \quad (5.5)$$

or

$$\beta = \left(\frac{1}{\lambda} + H^T H\right)^{-1} H^T T \quad (5.6)$$

and the corresponding output function of ELM is

$$f(x) = h(x)\beta = h(x)\left(\frac{1}{\lambda} + H^T H\right)^{-1} H^T T \quad \mathbf{(5.7)}$$

5.4 Conclusion

This chapter has expanded on the second algorithm used in the study. It has provided the concept behind the usage of the method and the limits within it. However, its application in the construction industry has not been discussed as it is rarely used by construction professionals. It is all the more reason why this investigation is being undertaken, namely to ascertain its efficacy against the multiple linear regression that is commonly used. Some of the strengths and limitations have also been discussed. This prepared the stage for its usage in the main modelling work. Data collected was used to develop predicting models through the application of the extreme learning machine.

Chapter 6: Results and Analysis

6.1 Introduction

The aim of this chapter is to present the empirical results. The main aim of this study was to investigate prediction methods for the completion cost of construction projects. The focus was on the capability of the multiple linear regression model and extreme learning machine model. This was empirically done by evaluating and comparing their performances in predicting construction completion cost. The performance measures being investigated were the overall accuracy on the training data set and the overall accuracy on the testing data set. In this chapter, results obtained from the two methods are presented and discussed. These results consist of the best performance models for both methods.

6.2 Demographics

There were a total of 911 contracts used in the study. Table 6.1 shows the demographic distribution of the data. The data that were used for the thesis were collected from all ten regions in Ghana. From the available data, 271completed projects representing 29.7% of the total sample was collected from Region 1; 148 representing 16.2% were collected from the second region; 35 representing 3.8% were from Region 3; 27 representing 3% were from Region 4, 119 representing 13.1% were from Region 5: 62 representing 6.8% were from Region 6; 38 representing 4.2% came form from Region 7; 16 representing 1.8% were from Region 8; 127 representing 13.9% were collected from Region 9, while 68 representing 7.5% were collected from Region 10.

There are three tendering types recorded in the data. These comprise competitive, quotation and sole source tendering. From the available data, 867 completed projects, representing 95.2% were the competitive tendering type; 43 representing 4.7% were quotation, while one (1) representing 0.1% was sole source tendering type. It is evident from the data that competitive tendering is the most dominant while sole source is the least. There were five different funders for the sampled projects. Out of the five, foreign funding (25) representing 2.7% of the total number of sample; 541 completed projects, representing 59.4% were funded by the GETfund; 214 representing 23.5% were funded by the Government of Ghana; 76 representing 8.3% were funded using internally generated funds (IGF); while non-governmental organizations (NGOs) funded 55, which represents 6%. From the available data, 550 representing 60.4% of the total samples had a scope change while 361 representing 39.6% had no scope change.

There were 665 contractors representing 73% who were classified as D1K1; 200 representing 22% were classified as D2K2; 33 representing 3.6% were classified as D3K3; while 13 representing 1.4% were D4K4. From table 6.2, a total of 426 of the total samples available representing 46.8% had a mobilization fund while 485 representing 53.2% did not have a mobilization fund. It is evident from the results that the majority of the contracts were not given mobilization. There were 109 contracts representing 12% that had liquidated damages while 802 contracts representing 88% did not have liquidated damages.

Table 6.1: Demographic analyses

Characteristics	Description	Frequency	Percentage
Region	Greater Accra	271	29.7
	Ashanti	148	16.2
	Brong Ahafo	35	3.8
	Central	27	3.0
	Eastern	119	13.1
	Northern	62	6.8
	Upper East	38	4.2
	Upper West	16	1.8
	Volta	127	13.9
	Western	68	7.5
Tendering type	Competitive	867	95.2
	Quotation	43	4.7
	Sole source	1	0.1
Funder	Foreign	25	2.7
	Getfund	541	59.4

	GOG	214	23.5
	IGF	76	8.3
	NGO	55	6.0
Scope change	Yes	361	39.6
	No	550	60.4
Provisional sum exceeded	No	470	51.6
	Yes	441	48.4

***Table 6.2*: Demographic analysis**

Characteristics	Description	Frequency	Percentage
Liquidated damages	No	802	88
	Yes	109	12
Mobilization fund	No	485	53.2
	Yes	426	46.8
Contractor type	D1K1	665	73
	D2K2	200	22
	D3K3	33	3.6
	D4K4	13	1.4

The data for the research was gathered from different consulting practitioners across the country who undertook projects for different years. In all, 911 completed projects which are located in all ten regions of Ghana were compiled and used for the analysis. Table 6.3 presents a breakdown of the projects used from each of the ten regions and their respective years. From table 6.3, it is evident that the majority of the projects were undertaken in the Greater Accra region in 2011 with the least number of projects being in the Upper West region. Again, the majority of the projects were undertaken in Greater Accra in 2012, 2013, 2014, 2015 and 2016. The Ashanti and Volta regions also recorded a high number of projects across the years. The least number of projects were undertaken in Upper West from 2011 to 2016 according to the data available in the study.

***Table 6.3*: Breakdown of projects across the ten regions**

	2011	2012	2013	2014	2015	2016
Greater Accra	26	91	16	25	34	79
Ashanti	9	58	6	15	30	30
Brong Ahafo	5	15	0	4	3	8
Central	4	15	1	0	1	6
Eastern	10	50	3	9	12	35
Northern	6	27	1	6	4	18
Upper east	7	9	0	1	8	13
Upper West	1	7	0	2	1	5
Volta	10	47	4	6	16	44
Western	6	27	0	4	5	26

6.3 Analysis of the sizes of cost overruns

The distribution of cost overruns which were experienced within all the projects in the research are classified and presented in Figure 6.1. A total of 35 projects (3.8%) of the total number of projects were completed within their initial budgeted cost. The majority of those projects which were completed within the initial budget cost were awarded through competitive bidding and did not also have scope change. Only a small proportion had scope change yet were able to complete within the budget. Only two projects were overestimated (under-runs). From the total number of projects that overran their initial budget, two (2) (0.2%), 55(6%), 811 (89%), 38 (4.2%) and five (5) (0.5%) experienced cost overrun in the categories less than zero, 0-5, 6-20, 21-50 and more than 50 respectively.

A total of 96.2% of the total projects in the study experienced cost overruns with the majority of the overruns in the 6-20 category. There were more overruns in the list of projects used in the study than underruns. From the data available, 619 (76.3%) of the number of projects in the 6-20 overrun category had D1K1 contractor types while there were three (3) (1.4%) with D4K4 contractor types. The large number of D1K1 contractor types in the 6-20 category could be attributed to experience and working capital as possible reasons. Larger sized firms with experienced contractor types are generally assumed to have a better cost performance compared to others. Table 6.4 presents the project statistics. The minimum and maximum cost overruns are -87.49% and 9900% respectively, with a very large range. The average cost overrun of all the projects analysed is 23.66%.

From the results obtained from the available data, there is a statistically significant difference among the contractor types regarding cost overrun. The analysis of variance (ANOVA) table for the difference among the contractor types is shown in Table 6.6. The p-value (0.00000494) for the ANOVA test shows that there is a statistically significant difference among the contractor types regarding cost overruns. The results of a Tukey test revealed that there were statistically significant differences in cost overruns between D1K1 and D3K3 contractor types (p-value=0.0000012), between D2K2 and D3K3 (p-value=0.0000046), and between D3K3 and D4K4 (p-value=0.0196). The results also indicated that there were no statistically significant differences in cost overruns between D1K1 and D2K2 (p-value=0.99), D1K1 and D4K4 (p-value=0.99), D2K2 and D4K4 (p-value=0.99).

***Table 6.4*: Breakdown of projects the ten regions**

Description	Statistics
Number of projects	911
Mean	23.66
Minimum	-87.49
Maximum	9900.00
Standard deviation	329.01

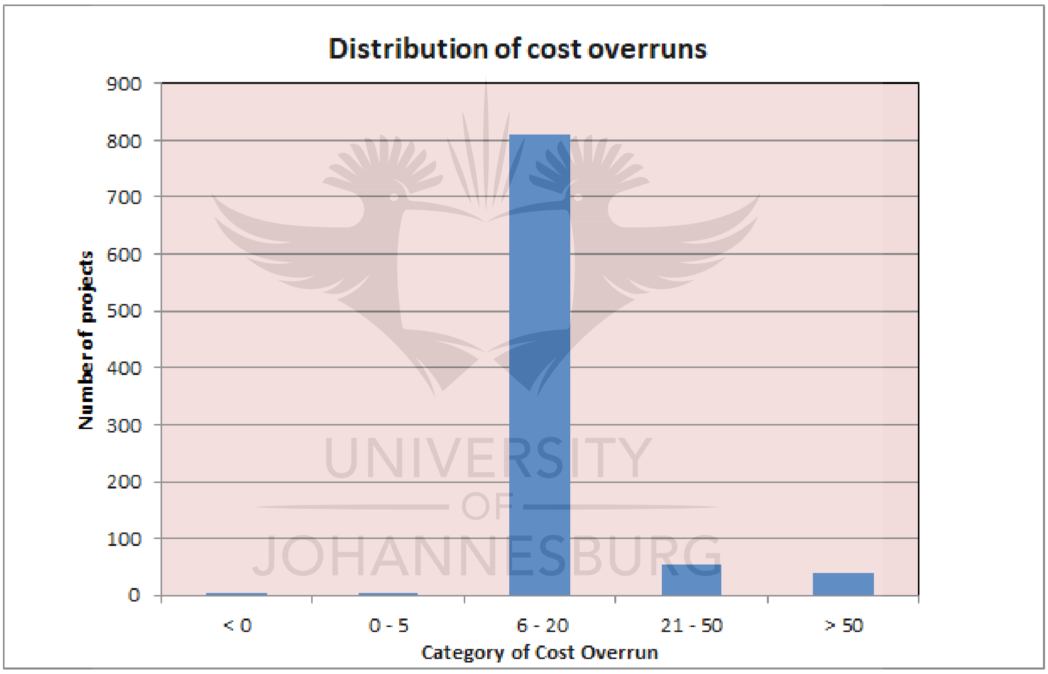


Figure 6:1 Distribution of cost overruns

***Table 6.5*: ANOVA test for variation in cost overrun among contractor types**

	Df	SS	MS	F value	P-value
Contractor type	3	2925808	975269	9.255	0.00000494
residuals	907	95579328	105380		

6.4 Proposed multiple linear regression models

In an effort to predict project completion cost, one of the most commonly used methods is the multiple linear regression. It is a method generally used to predict quantitative response variables given a number of predictor variables. Three models that seek to use the predictor variables to predict the completion cost of the projects are proposed.

6.4.1 Regression Model A

The first model (named Regression Model A) considers all the predictor variables that are available in the data set.

6.4.1.1 Descriptive statistics

Descriptive statistics were used to make simple descriptions and comparison of the numeric predictor variables used. The descriptive statistics are presented in Table 6.6. From Table 6.6, it can be seen that the minimum and maximum duration for the projects used in the study were one (1) and 24 months respectively, while the average duration for the projects was nine (9) months. The minimum gross floor area was 12meters while the maximum gross area was 8902. On average, the gross floor area was 997.9. The minimum and maximum number of storeys were 1 and 6 respectively, while the average number of storeys is 1.5. The minimum contract sum for the projects in the study was 15841 cedis, the maximum was 15969569 and the average contract sum was 1408219 cedis. The minimum completion cost for the projects in the study was 17425 cedis, and the maximum was 16847896 cedis. The minimum and maximum completion costs were 17425 cedis and 16847896 cedis respectively. The average completion cost was 1561301 cedis. The minimum and maximum consultant budget costs were 15128 cedis and 15809874 cedis respectively. The average consultant budget cost was 1434666 cedis.

***Table 6.6*: Breakdown of projects across the ten regions**

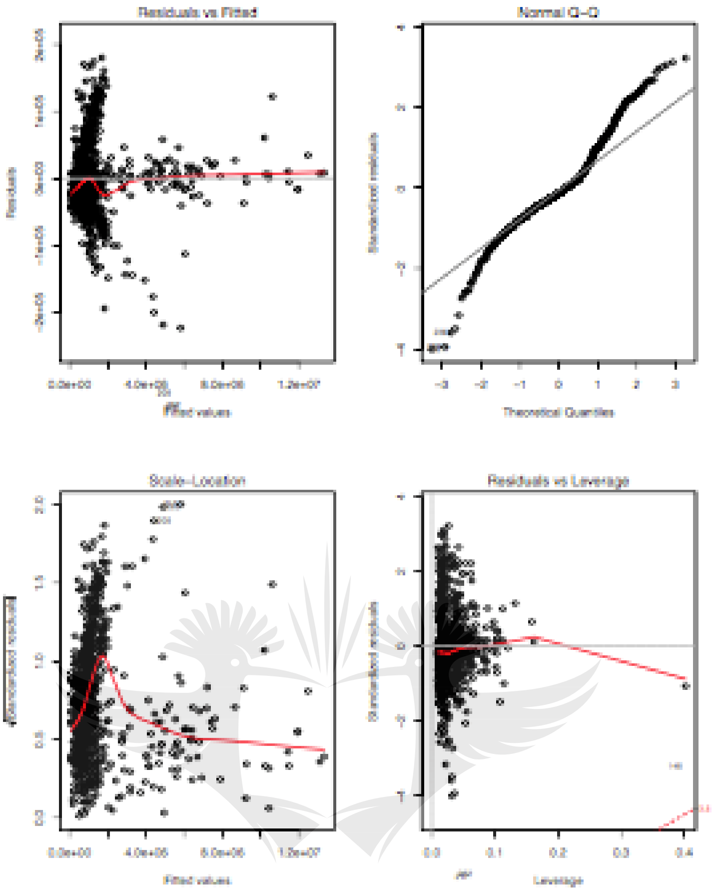
Variable	Minimum	Maximum	Mean	Std. dv
Initial duration	1	24	9.081	4.248467
Gross floor area	12	8902	959.5	894.0633
Number of storeys	1	6	1.491	0.9426858
Initial contract	15841	12060853	1305561	1656508
Completion cost	17425	13327243	1448531	1819809

Table 6.7 shows a summary of the coefficients for the predictor variables in the final model. The F-statistics are 17210 with a p-value less than 0.000. The adjusted R square is 99.82%, which is an indication that the predictor variables explain 99.82% of the variability in completion cost of projects. The Durbin Watson statistics are 1.9 with a p value of 0.466. The non-significant p-value suggests a lack of autocorrelation, and conversely an independence of the errors. The lag value of one indicates that each observation is being compared with the one next to it in the dataset. A significant p-value for the non-constance variance ($p - \text{value} < 0.000$) and for the Breusch Pagan test ($p - \text{value} < 0.000$) suggests heteroscedasticity. The global test of linear model assumptions also indicated that the data did not meet all the statistical assumptions of multiple linear regression. All the predictor variables have a variance inflation factor less than five (5), indicating no issues of multicollinearity.

***Table 6.7*: Regression results**

Variable	Estimate	t value	p value
Intercept	41580	2.908	0.003727
Gross floor area	4.889	1.113	0.265850
Number of storeys	5120	1.407	0.159800
Ashanti	9194	1.560	0.1119060
Brong Ahafo	5335	0.498	0.618562
Central	-17370	-1.430	0.153118
Eastern	-3927	-0.613	0.540161
Northern	-6593	-0.764	0.445062
Upper East	11740	1.163	0.245266
Upper West	11330	0.691	0.489832
Volta	8012	1.264	0.206584
Western	-3884	-0.495	0.620947
Initial	1.098	447.773	<0.0001
GETFUNG	-30830	-2.535	0.011428
GOG	-35880	-2.864	0.004291
IGF	-6094	-0.409	0.682543
NGO	-20640	-1438	0.150767
Initial duration	-3419	-5.229	<0.0001
Scope change	13930	3.055	0.002324
Provisional sum exceeded	16570	3.716	0.000215
Liquidated damages	-8519	-1.422	0.155478
Mobilization fund	9983	2.524	0.011778
Quotation	-40390	-3.374	0.000776
Sole Source	-51930	-0.906	0.365098
D2K2	7431	1.559	0.119312
D3K3	-18450	-1.660	0.097254
D4K4	-23670	-1.347	0.178436

Figure 6.2: Regression output.



From Figure 6.2, the normal Q-Q plot suggests that the residual values are slightly skewed. There seems to be a pattern in the plot of the residuals and the fitted values, which indicates that the error terms do not have a constant variance. The points in the scale-location graph do not seem random which is also an indication that the constant variance assumption has not been met. There are still outliers as evident in the residual versus leverage plot.

Figure 6.3: Regression output.

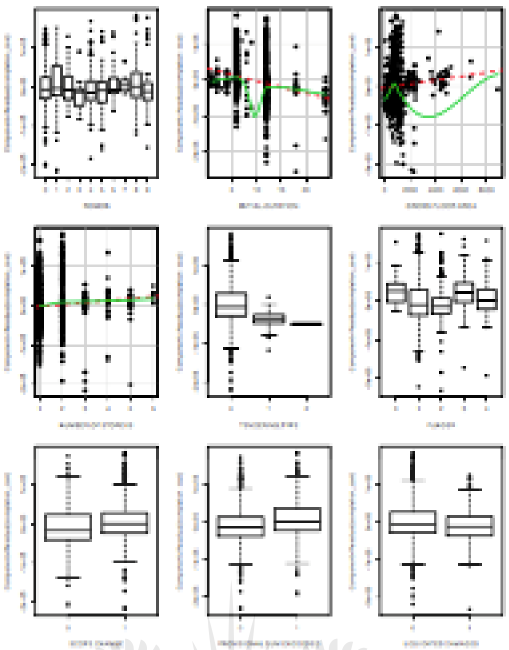


***Table 6.8*: Relative importance of the predictor variables**

Variable	Relative importance
Region	0.0024154548
Tendering type	0.0010275153
Funder	0.0071957016
Contractor type	0.00170012499
Initial duration	0.1078397744
Gross floor area	0.2196867313
Number of storeys	0.1856847302
Scope change	0.0004507292
Provisional sum exceeded	0.0004588221
Liquidated damages	0.0003163992
Mobilisation fund	0.0003554816
Initial contract	0.4719266474

Table 6.8 presents the relative importance of each of the predictor variables in the model. The relative importance list indicates each variables contribution to predicting the completion cost of the project. From the results, initial contract sum was the most important predictor variable followed by gross floor area and number of storeys. Liquidated damages were the least important variable.

6.4.2 Regression Model B

The second model (named Regression Model B) uses variables that are only statistically significant. This variable selection approach is intended to select the best subset of predictor variables that are related to the completion cost of the projects using a statistical approach. Forward selection, backward selection and a combination of both forward and backward selection methods was used to select statistically significant variables. The selected predictor variables are used in regression model B to predict completion cost.

***Table 6.9*: Regression results**

Variable	Estimate	t value	p value
Intercept	41520	3.008	0.002710
Initial duration	-3236	-5.114	<0.0001
Number of storeys	6123	1.726	0.084627
Quotation	-41380	-3.470	0.000546
Sole Source	-60560	-1.064	0.287585
GETFUNG	-28560	-2.361	0.018460
GOG	-35230	-2.824	0.004848
IGF	-4857	-0.327	0.743380
NGO	-21600	-1.511	0.131130
Scope change	13060	2.896	0.003880
Provisional sum exceeded	16700	3.813	0.000147
Mobilization fund	8415	2.166	0.030568
D2K2	7883	1.662	0.096860
D3K3	-16730	-1.531	0.126017
D4K4	-25740	-1.478	0.139805
Initial contract	1.100	539.413	<0.0001

The model above was generated after stepwise regression was performed to select predictor variables that were important. The F-statistics are 59870 with a p-value less than 0.000. The adjusted R square is 99.9%, which is an indication that the predictor variables explain 99.9% of the variability in completion cost of projects. The Durbin Watson statistics are significant, indicating the presence of autocorrelation. A significant p-value for the non-constance variance (p-value< 0.000) and for the Breusch Pagan test (p-value¡0.000) suggests heteroscedasticity. The global test of linear model assumptions also indicated that the data did not meet all the statistical assumptions of multiple linear regression. All the predictor variables have a variance inflation factor less than five (5), indicating no issues of multicollinearity.

***Table 6.10*: Relative importance of the predictor variables**

Variable	Relative importance
Tendering type	0.0012499348
Funder	0.0090826689
Contractor type	0.0021906925
Initial duration	0.1439855080
Number of storeys	0.2493825205
Scope change	0.0005802311
Provisional sum exceeded	0.0005560175
Mobilisation fund	0.0004342095
Initial contract	0.5915825554

The relative importance list of the predictor variables shows that initial contract sum was the most important variable. Also, the number of storeys and the initial duration of the project contributed to the prediction of the completion cost of the project. The least important variable was the mobilisation fund.

Figure 6.4: Regression output from model 2.

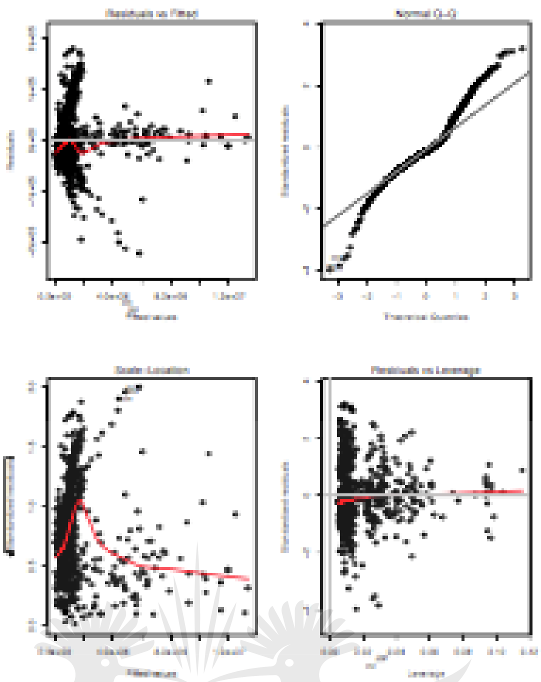
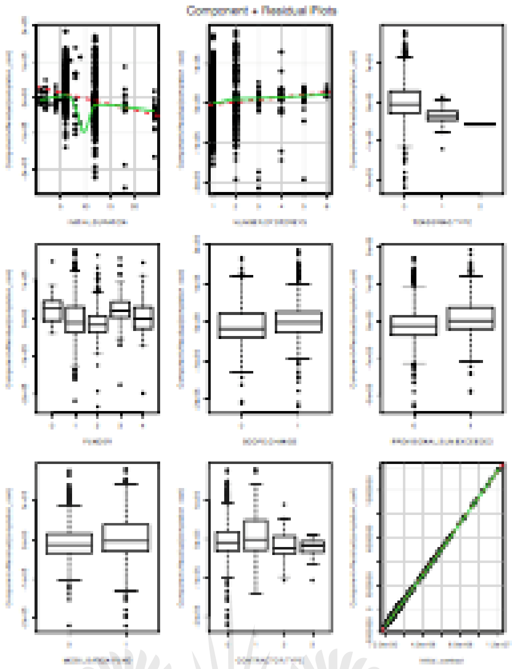


Figure 6.5: Regression output from model 2.



6.4.3 Regression Model C

The third model (Regression Model C) uses only statistically significant predictor variables from regression Model A. The F-statistics are 57250 with a p-value less than 0.000. The adjusted R square is 99.9%, which is an indication that the predictor variables explain 99.9% of the variability in the completion cost of projects. The Durbin Watson statistics are significant, indicating the presence of autocorrelation. A non-significant p-value for the non-constance variance (p-value=0.05) suggests that the error variance is constant. The global test of linear model assumptions also indicated that the data did not meet all the statistical assumptions of multiple linear regression. All the predictor variables have a variance inflation factor less than five (5), indicating no issues of multicollinearity.

***Table 6.11*: Regression results**

Variable	Estimate	t value	p value
Intercept	17460	2.582	0.009976
Initial contract	-1100	551.417	<0.0001
Initial duration	-3016	-4.943	<0.0001
Scope change	12000	2.667	0.0007805
Provisional sum exceeded	16970	3.847	0.000129
Mobilization fund	8607	2.200	0.028097
Quotation	-2.1160	-2.256	0.024293
Sole Source	-58220	-1.015	0.310567
D2K2	8360	1.706	0.088383
D3K3	14320	-1.308	0.191080
D4K4	-26450	-1.508	0.131914

The relative importance list of the predictor variables shows that initial contract sum was the most important variable. Also, the initial duration of the project contributed to the prediction of the completion cost of the project. The least important variable was the mobilisation fund.

***Table 6.12*: Relative importance of the predictor variables**

Variable	Relative importance
Tendering type	0.0018686198
Contractor type	0.0026211833
Initial duration	0.2090034542
Scope change	0.0009134257
Provisional sum exceeded	0.0006530319
Mobilisation fund	0.0005962679
Initial contract	0.7833657562

6.5 Model selection

It is important to consider selecting the best model by looking at the simplicity of the model. The chosen model should explain much of the variation within the data. The three multiple linear regression models considered are as follows:

• Model A: all predictor variables;

• Model B: predictor variables selected using stepwise regression; and

• Model C: using only statistically significant predictor variables from Model 1.

In Model A, all the predictor variables that were identified and available in the data set were used to build the model. In Model B, the predictor variables were selected using a statistical approach, stepwise regression and the variables that were selected to build the model. In Model C, only the statistically significant variables in model A were used. Four selection techniques, namely coefficient of determination (adjusted R^2), Akaike information criteria (AIC), Bayesian information criteria and analysis of variance (ANOVA) were used to select the most statistically significant model.

***Table 6.13*: Comparison of actual and predicted completion costs using regression**

	Regression model					
Actual	**Model 1**	**PE**	**Model 2**	**PE**	**Model 3**	**PE**
572826.5	**562966.6**	**1.721273**	**574896.7**	**-0.3614**	**546462.6**	**4.602421**
640636.9	**651143.2**	**-1.63999**	**641307.7**	**-0.10472**	**614392.9**	**4.096539**
584571.7	**561005.7**	**4.03132**	**559838.3**	**4.231023**	**554199.8**	**5.195575**
1002446	**989067.3**	**1.334573**	**986427.6**	**1.597894**	**959295.5**	**4.304487**
868129.7	**813262.3**	**6.320191**	**811219.9**	**6.555449**	**818743.7**	**5.688785**

1147136	**1082153**	**5.664806**	**1071275**	**6.613156**	**1062332.2**	**7.392691**
557715.4	**655407.4**	**-17.5165**	**644376.5**	**-15.5386**	**651701.8**	**-16.852**
1737770	**1606426**	**7.558196**	**1612771**	**7.193056**	**1617777.8**	**6.04957**
1015939	**944662.6**	**7.015791**	**950520.2**	**6.439221**	**959710.3**	**5.53463**
1735593	**1611484**	**7.150828**	**1604392**	**7.559411**	**1603774.3**	**7.595018**

From the three models, the (adjusted R^2) for each of the three models were very similar as shown in the Table 6.14. From the results obtained, the best model is Model B based on the values of the AIC and BIC. The AIC and BIC for model B are smaller than the AIC and BIC for the other two models. Also, from the ANOVA result, model B was the best model indicating that the extra predictor variables added to models A and C do not add extra linear prediction.

***Table 6.14*: Model selection**

	AIC	BIC	Adjusted R^2
Model A	21665.35	21799.03	99.9
Model B	21657.10	21724	99.9
Model C	21667.55	21738	99.9

6.6 Extreme learning machine results

The ELM technique was used on the same data set used in the multiple linear regression. The training set and testing set are randomly generated from the whole data set before the simulation. Different combinations of the regularisation coefficient and kernel parameter were used to estimate the generalised accuracy. The kernel type which was used in the model is the radial basis function (RBF). The training speeds of the ELM as well as the testing speed for each of the parameters are also listed. The results of the predictions using each of the models are in shown in Table 6.12. The regularisation coefficient is given in brackets. As the regularisation coefficient increases, the accuracy of the model increases. This is shown in Table 6.12. The percentage errors (PE) of each of the models are also predicted. It is evident from the result that, as the number of regularisation coefficients increases; the PE then decreases, leading to a better prediction. The model with the highest accuracy is Model 7, which has 700 as the regularisation coefficient.

Figure 6.6: Plot of actual and predicted values

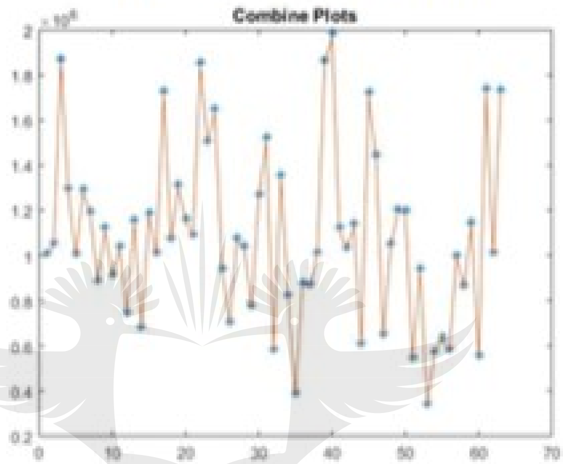


Figure 6.6 shows the deviations between actual and predicted completion costs. The asterisk in Figure 6.6 represents the actual values from the testing data while the graph represents the predicted values from the model using ELM. From the graph, it is evident that the predicted values are closer to the actual values, which is indicative of how well the model performs.

***Table 6.15*: Comparison of actual and predicted completion cost using ELM**

	Extreme Learning Machine									
Actual	Model 1 (100)	PE	Model 3 (200)	PE	Model 5 (300)	PE	Model 6 (350)	PE	Model 7 (700)	PE
572826.5	567155	0.990099	569976.6	0.497512	570923.4	0.332226	571194.5	0.2849	572009.4	0.142653
640636.9	634293.9	0.9901	637449.6	0.497513	638508.5	0.332227	638811.7	0.284901	639723	0.142654
584571.7	578783.8	0.990099	581663.4	0.497512	582629.6	0.332226	582906.2	0.2849	583737.8	0.142653
1002446	992520.5	0.990096	997458.4	0.497509	999115.3	0.332223	999589.7	0.284897	1001016	0.14265
868129.7	859534.4	0.990099	863810.7	0.497512	865245.6	0.332226	865656.4	0.2849	866891.3	0.14653
1147136	1135779	0.990096	1141429	0.497509	1143325	0.332223	1143868	0.284897	1145500	0.14265
557715.4	552193.4	0.990099	554940.6	0.497513	555862.5	0.332226	556126.4	0.284901	556919.8	0.142654
1737770	1720564	0.990097	1729124	0.497511	1731997	0.332224	1732819	0.284898	1735291	0.142652
1015939	1005880	0.990099	1010884	0.497513	1012563	0.332226	1013044	0.284901	1014489	0.142654
1735593	1718409	0.990097	1726958	0.49751	1729827	0.332224	1730648	0.284898	1733117	0.142651

***Table 6.16*: T-test for the Best performing model for Regression and ELM**

Best Regression Model	Best ELM Model
574896.7	572009.4
641307.7	639723
559838.3	583737.8
986427.6	1001016
811219.9	866891.3
1071275	1145500
644376.5	556919.8
1612771	1735291
950520.2	1014489
1604392	1733117
t-test	**0.837326008**

6.7 Findings

From the analyses the following were found out;

Competitive tendering is the dominant method of procurement in the building sector construction projects delivery in Ghana. Scope change is prevalent in more than 50% of the completed projects analysed. Projects completed within budgeted cost are mostly awarded through competitive tendering and had no scope change during delivery. The average cost overrun on projects is 23.6 %. Comparatively the Extreme learning machine model outperforms the best multiple linear regression model. Even though is not evident from the significance of the t-test (p-value=0.837), but it can be seen from the percentage error values that the predictions are better using the ELM than MLR. Thus ELM proves better in accuracy than the MLR.

6.8 Conclusion

Countless information is generated during the execution of construction projects. When the information generated is properly documented and stored for use, much can be done with it. The information can be of importance to all stakeholders when extracting useful aspect of it for other analysis but it must be done during delivery of the projects in a manner that it can be easily re-used.

Using multiple linear regressions and the extreme learning machine, cost predicting models were developed (and compared) with 911 construction projects that were executed from 2011 to 2016. The extreme learning machine achieved a more a robust and improved result. When compared to the models developed using multiple learning regression. The output of the model, especially the one developed using the extreme learning machine, suggests a promising future for cost estimating. It shows that improved data storage could enhance the predicting accuracy of construction cost. The regularisation coefficient of the extreme learning machine could further leave room for generating alternatives and establishing a range for cost estimates. The strategy used in this work could be adapted to other estimating practices where data is available and can be extended.

Chapter 7: Discussion

7.1. Introduction

Time, quality and value are the cardinal concerns of any clients in the construction industry, especially when they are about to put up a facility. Early cost estimates enable construction clients to undertake cost feasibilities, look for funding, examine bids and they serve as a guide in monitoring, evaluating and controlling cost during project execution phases. To a larger extent, lifecycle costings are done to ascertain the economic viability of projects, and decisions taken at the preliminary stages of the project delivery have far-reaching effects (economically) and can determine whether the project sees the light of day.

Unfortunately, construction infrastructure deliveries predominately overrun their cost estimates, thus displeasing owners, funding partners, consultants, contractor and citizens at large. Investors sometimes face the problem of not obtaining a profit on their investment for long periods of time. Consultants and contractors suffer image crises, liquidity problems and cash flow irregularities and in the cases of governmental projects, the citizens have a bad perception of the initiators, regulators, and executing entities as well.

To this end, this research has addressed the issue of cost overrun in construction projects (particularly in building projects) by looking at what the issues are, why these are prevalent and how to improve on addressing the issue. The study aligns the arguments and findings contained in this thesis with the overall aim and objectives outlined in the preliminary chapter of the work. This concluding part of the study answers the objectives and questions posed and reveals the theoretical and practical inferences that are made of the research. Guidance for project delivery and management, especially in the building construction sector, was suggested and further research possibilities were enumerated.

7.2. Appraisal of original aim and objectives

The overall aim of the current study was to develop an alternative improved prediction cost estimating model for public construction projects in Ghana and to identify the contributing factors that collectively influence cost overrun in the public sector projects.

Consequently, to achieve this aim, the study used a mixed method by undertaking a literature review, conducting an explorative interview with experienced practitioners in the construction industry of Ghana, and mining historical data to unveil the contributing factors that influence cost overruns. These factors, together with completed projects' data, were analysed using multiple regression and extreme learning machine to ascertain which of them would give an improved accuracy model. The conclusion of the study is presented in line with the objectives in the following sections.

The first objective was to propose appropriate cost overrun influence factors in construction projects in Ghana. In achieving this objective, a review of relevant literature was undertaken, an explorative interview with seasoned practitioners was conducted, and historical data was mined, revealing the factors that contribute to cost overrun. It came to light that cost overrun factors are not straightforward: rather, they are multi-faceted in nature and run across the phases of project delivery. Further, more of the cost overrun research carried out was mostly conducted in highway- or road-related construction as against building works. It was also revealed that the contexts in which projects are delivered contribute to some factors. In all, with the aid of an explorative interview with experienced practitioners and historical data mining 12 predictive variables namely; gross floor area, award contract sum, tendering type, contractor type, Liquidated damages, Advance mobilisation, region (location), Funding partners, project duration, Number of floors, scope change and provisional sum were uncovered and deemed fit for the models developed. Findings from the literature show the need for more effort to be put into research aimed at improving the accuracy of modelling cost estimates.

the initial contract sum, tendering type, scope change, project duration, funding agency, region (location), mobilization fund availability, gross floor area, number of storeys, provisional lump sums, liquidated damages and contractor type.

The second objective was to investigate the patterns and trends of cost overruns in the construction industry of Ghana. In this regard projects were drawn from the 10 regions of Ghana in order to examine the patterns' and trends' regional characteristics in terms of cost overruns as shown in the results and analysis. The industry activities were reviewed, revealing the funding agencies, procurement route, ceiling's and executing entities set up and how they operate. The effect of funding partners, procurement options, and contractors was analysed in relation to cost overrun to reveal the trends and patterns in the model development.

The third objective was to identify factors that affect cost overrun of the construction industry in Ghana and select those that are statistically significant. Arriving at the factors as discussed, multiple regression (using stepwise regression) was used in analysing and selecting the statistically significant ones as discussed Chapter Six encompassing results and analysis. Furthermore, collecting and extracting 911 data set of completed projects across the 10 regions of Ghana (mainly education-related projects) for the model and the relative importance of the resulting variables were ascertained using both multiple regression and extreme learning machine development.

The fourth, objective was to use the data set acquired to develop models using multiple regression and extreme learning machine. It was also an objective to compare the developed models in order to discover the predicting accuracy improvement of the two methods (multiple regression and extreme learning machine). To achieve these objectives, the theory and process of using multiple regressions and extreme learning for constructing models were used for 911 completed projects. Of the data set collected, 80% were used in training the cost-estimating models. The predictive variables for the model in this exercise include the initial contract sum, tendering type, scope change, project duration, funding agency, region (location), mobilization fund availability, gross floor area, number of storeys, provisional lump sums, liquidated damages and contractor type. In the modelling process of the two methods, three operations were carried out: firstly, all variables were used to model; secondly, stepwise regression was done to model; and lastly, only significant variables were used to develop the predictive model.

After achieving successful results, the remaining 20% was used to test models. The two different modelling methods were used sequentially with the aim of discovering the more improved accurate and appropriate model that will transform completed projects information from historical data into cost models, thereby aiding the process of estimating construction cost. Details of the results and analysis carried out using the two methods were discussed in Chapter Six. The testing results have been presented in tables 6.13 and 6.12. The results show that multiple regression models achieved a predicting accuracy that is less superior to the result achieved by extreme learning machine. This is not evident in table 6.16 since the t-test value is 0.837326008. However, the percentage error of ELM is far lower than that of MLR. Thus the percentage errors show significance in the variance of prediction accuracy

The construction clients are ever looking for to an improvement in the early cost estimates and information available from historical data can aid the improvement through modelling. Construction entities usually keep information on completed projects and data from this information can be converted into decision-making tools for improvement in cost estimating. Data mining techniques for digging coherent patterns and reliable relationships between factors were used in this study to develop a predicting completion cost estimate model using data collated from the construction industry in Ghana. Extreme learning machine and multiple regressions were used on the data to develop a predicting model in estimating the final or completion cost when construction is yet to begin. The performance of the extreme learning machine and multiple regressions were shown in table 16.12. The extreme learning machine model far outweighs the multiple regressions models as depicted in table 6.13 and thus shows an improvement in cost estimating, thereby aiding attempts to reduce overrun in construction cost.

7.3 Contributions and worth of the research

The aim and objectives of this research were achieved as shown from literature, methodology and the practical level of its findings. However, it is important that the contribution from the results and findings is deduced and placed within the larger context of construction project delivery research and practice.

7.3.1 Theoretical contribution

This research contributes to existing knowledge in construction projects cost prediction or modelling advances. It reveals the use of extreme learning machine with multiple regressions for developing predicting cost-estimating models built on historical data from completed construction projects to determine the best predictive model among the developed models. In the field of construction final cost estimating, cost overrun estimating and related research, the use of extreme learning machine, combined with multiple regressions and deducing the best predictive model among the two methods has never been done in any research before as far as can be established.

7.3.2 Practical contribution and worth

The use of extreme learning machine and multiple regressions for modelling construction cost has shown a potential possibility for converting historical data that exist within the construction entities into decision-making kits; more importantly where inadequate or lack of information abounds, the predictive models developed would be useful to clients, consultants, contractors, ministries and funding partners at the contract award stage to

- undertake prior projections or prediction of variances in cost estimates;
- enhance to some degree of certainty initial cost decisions taken;
- forecast potential cash flow-related problems sufficiently early;
- plan towards securing adequate funds for project delivery; and
- reduce public outcry regarding overruns in the industry, thereby improving public confidence.

The model developed can be of direct benefit to industry players as it can be converted to desktop software for predicting cost. Companies can profile their projects to compile homogenous data for use in predicting. The model will help to improve the planning of projects, and avoid the postponement or neglect of projects

7.4 Recommendations

The following are recommendation stemming from the results, findings and analysis of the study:

- Any reliable model that is developed depends on good quality data and the use of data can be valuable to any existing organization if taken advantage of. The construction industry therefore needs to recognize the importance of creating a proper data warehousing system and using the large amount of data accessible to them from completed projects to aid the estimating process.
- Project management process must include relays of data and data gathering techniques to enable continuum in the project delivery process so as not to lose valuable information that could aid sound construction cost decision-making and be helpful in predicting cost for future projects.

7.5 Limitation of the research

The predictive model developed in this study was tested using historical data set aside for testing. To enhance the validation of the model, testing is needed to validate the predictive model by using a project at the award stage. This will lead to monitoring till the project is completed to ascertain the degree of certainty of the actual completed cost and predicted cost. Time schedule and scope of the current study could not allow for this validation to be carried out.

It is possible that a thorough breakdown of building projects into more robust detail beyond what has been used could produce more relevant information for the models to be further improved.

It is further envisaged that improved data warehousing in the construction industry could lead to an enhanced predictive model where these models were developed based on the existing data available in the industry.

7.6 Further research

Despite useful and important outcomes having been obtained from the study, there is more room for further study and enhancement. An extended study in the area of extreme learning, comparing results using different kernel types and optimal combinations of regularization coefficients and kernel parameters could be explored. Other deep learning methods such as recurrent neural network could be used. The use of support vector machines is also encouraged and recommended for further research.

References

Addo-Abedi, F.Y. (1999). Sustained Development of the Local Contracting Industry in the Developing Countries. *Proceedings of the Second International Conference on Construction Industry development*, October. pp. 27 – 29

Adeli H. and Wu M. (1998). Regularization Neural Network for Construction Cost Estimating., *Journal of Construction Engineering and Management* .124(1).

Adeli, H. (2001). Neural networks in civil engineering: 1989-2000. *Computer-Aided Civil and Infrastructure Engineering*, **16**(2), pp.126- 42.

Adeli, H and Yeh, C. (1989). Perceptron learning in engineering design. *Computer-Aided Civil and Infrastructure Engineering*, **4**(4), pp. 247- 56.

Ahiaga-Dagbui, D. D. and Smith, S.D. (2014). Exploring escalation of commitment in construction project management: Case study of the Scottish Parliament project. In: *Procs 30th Annual ARCOM Conference*, Raiden, A, Ed., Portsmouth, UK: Association of Researchers in Construction Management

Ahiaga-Dagbui, D.D. and Smith, S.D. (2013). "'My cost runneth over': data mining to reduce construction cost overruns", in Smith, S.D. and Ahiaga-Dagbui, D.D. (Eds), Proceedings of 29th Annual ARCOM Conference, Association of Researchers in Construction Management, Reading, MA, pp. 559-568.

Ahiaga-Dagbui, D.D. and Smith, S.D. (2014). "Dealing with construction cost overruns using data mining", *Construction Management and Economics*, Vol. 32 Nos 7/8, pp. 682-694.

Aibinu, A. A. and Pasco, T. (2008). The accuracy of pre-tender building cost estimates in Australia. *Construction Management and Economics*, **26** (12), pp. 1257 - 69

Aibinu, A. and Jagboro, G. (2002). The effects of Construction Delays on Project Delivery in Nigerian Construction Industry. International Journal of project

Management, 20(8), pp. 593 - 599.

Akintoye, A. and Skitmore, M. (1994). Models of UK Private Sector Quarterly Construction Demand. *Journal of construction management and economics, issue 12*, pp. 3 – 13

Akintoye, A. S. and MacLeod, M. J. (1997). Risk analysis and management in construction. *International Journal of Project Management*, **15**(1), pp.31-8.

Akintoye, A. (2000). Analysis of factors influencing project cost estimating practice. Construction management and economics, 18(1), 7789.

Akintoye, A. and Fitzgerald, E. (2000). A survey of current cost estimating practices in the UK. Construction management and economics, 18(2), 161172.

Akusok, A., Björk, K.M., Miche, Y. and Lendasse, A. (2015). High performance extreme learning machines: a complete toolbox for big data applications. IEEE Access, 3, pp.1011-1025.

Al-Momani, A.H. (1996). "Construction cost prediction for public school buildings in Jordan", Construction Management and Economics, Vol. 14 No. 4, pp. 311-317.

Alex D. P., Hussein M. A., Bouferguene, A. and Fernando P.E. (2010). Artificial Neural Network Model for Cost Estimation: City of Edmonton's Water and Sewer Installation Services, J. Constr. Eng. Manage., 136(7): pp.745-756

Ameh, O., Soyingbe, A. and Odusami, K. (2010). Significant Factors Causing Cost Overruns in Telecommunication Projects in Nigeria. Journal of Construction in Developing Countries, 15(2), pp. 49 - 67.

Amnesty International Report. May 2009.

Amoah, P., Ahadzie, D. K. and Dansoh, A. (2011). The factors affecting construction performance in Ghana: the perspective of small-scale building contractors.

Amusan, L.M. (2011). covenantuniversity.edu.ng. [Online] Available at: http://webcache.googleusercontent.com/search?q=cache:kE8R_E7kWucJ:eprints.covenantuniversity.edu.ng/121/1/COST_PERFORMANCE2.doc+&cd=1&hl=en&ct=clnk&gl=za [Accessed 10 August 2017].

Ashworth A. (2008). Pre-Contract Studies Development Economics, *Tendering and Estimating*, 3rd edn, London, Blackwell, pp. 293-294.

Asiedu, R.O. and Alfen, H.W. (2014a). “Factors engendering cost misrepresentation of public sector projects in Ghana”, *International Journal of Sustainable Construction Engineering and Technology*, Vol. 5 No. 2, pp. 13-24.

Asiedu, R.O. and Alfen, H.W. (2014b). “Understanding the Critical Failure Factors Contributing to Cost Overruns in Public Sector Projects”, 4th International Conference on Infrastructure Development in Africa, Kwame Nkrumah University of Science and Technology, Kumasi, 26-28 March, pp.106-120.

Asiedu, R.O. and Alfen, H.W. (2015a), “Understanding the underlying reasons behind time overruns of government building projects in Ghana”, *KSCE Journal of Civil Engineering*, Vol. 20 No. 1, pp.1-9.

Asiedu, R.O. and Alfen, H.W. (2015c), “Assessing the frequency of factors engendering cost overruns using frequency indices and factor analysis for building projects in Ghana”, *Journal of International Real Estate and Construction Studies*, Vol. 3 No. 2, pp.1-22.

Attalla, M. and Hegazy, T. (2003). "Predicting Cost Deviation in Reconstruction Projects: Artificial Neural Networks versus Regression." *ASCE Journal of Construction Engineering and Management,* 129(4), pp. 405-411.

Ayeribi, D. (2005). Strategic planning practice of construction firms in Ghana. *The journal of construction management and economics, issue 23*, pp.163 -168.

Baccarini, D. (2005). Estimating project cost contingency – Beyond the 10% syndrome. In: *2005 Australian Institute of Project Management Conference*: AIPM

Baloi, D. and Prince, A. (2003). Modelling Global Risk Factors Affecting Consruction. Cost Performance, International Journal of Project Management, Volume 21, pp. 261 - 269.

Baloyi, L. and Bekker, M. (2010). Causes of construction cost and time overruns: The 2010 FIFA World Cup stadia in South Africa, Acta Structilia: 18(1)

Bartlett, P.L. (1998). The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. IEEE transactions on Information Theory, 44(2), pp.525-536.

Berenson, M., Levine, D., Szabat, K.A. and Krehbiel, T.C. (2012). *Basic business statistics: Concepts and applications*. Pearson Higher Education AU.

Bode, J. (2000). Artificial neural networks for cost estimation: simulations and pilot application. *International Journal of Production Research*, **38**(6), pp.1231-54.

Bordat, C., McCullouch, B.G., Sinha, K. C. and Labi, S. (2004). *An Analysis of Cost Overruns and Time Delays of INDOT Projects.*, Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, Publication FHWA/IN/JTRP-2004/07.

Boussabaine, A. and Elhag, T. (1997). A neurofuzzy model for predicting cost and duration of construction projects. *RICS Research (9 p.). The Royal Institution of Chartered Surveyors*.

Branke, J. (1995). Evolutionary algorithms for neural network design and training. *Proceedings of the First Nordic Workshop on Genetic Algorithms and its Applicatio*ns.

Bromilow, F. J. (1969). "Contract time performance expectations and the reality." *Building Forum*, 1(3).

Bromilow, F. J., Hinds, M. F. and Moody, N. F. (1980). "AIQS Survey of Building Contract Time Performance." *The Building Economist*, 19(2), pp.79-82.

Bryman, A. (2012). *Social research methods*. Oxford university press. Burger, R (2003) Contingency, quantifying the uncertainty. *Cost Engineering*, **45**(8), pp.12-7.

Burrows, T. K., Pegg, I., and Martin, J. (2005). "Predicting Building Construction Duration." *AACE International Transactions*, PS.14.11-PS.14-19.

Busby J.S. and Hughes, E.J. (2004). 'Projects, pathogens and incubation processes' *International Journal of Project Management*, vol 22, pp. 425-434.

Cantarelli, C.C., Molin, E.J., Van Wee, B. and Flyvbjerg, B. (2012). "Characteristics of cost overruns for Dutch transport infrastructure projects and the importance of the decision to build and project phases", Transport Policy, Vol. 22 No. 1, pp. 49-56.

Center for Affordable Housing Finance in Ghana, (2016). (online) available at http://housingfinanceafrica.org/countries/ghana/. Accessed on 29 December 2016)

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

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

Chen, S., Cowan, C.F. and Grant, P.M. (1991). Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on neural networks*, 2(2), pp.302-309.

Cheung, S. O., Yiu, K. T. W. and Chim, P. S. (2006). "*How relational are construction contracts?*" J. Prof. Issues Eng. Educ. Pract., 132(1), pp.48–56.

CIOB (2014). 5th edn *Code of Practice for Project Management for Construction and Development,* London, Blackwell.

Creedy, G. D. (2006). *Risk factors leading to cost overrun in the delivery of highway construction projects*, PhD, School of Urban Development, Faculty of Built Envornment and Engineering, Queensland University of Australia.

Creswell, J.W. (2009). Research design: Qualitative, quantitative, and mixed methods approaches. *Third ed. California, USA: Sage Publications, Inc.*

Crown Agents. (1998). The World Bank Procurement Audit in Ghana. Value for Money Audit Report for Ghana, Crown Agents Overseas Governments and Administrations Ltd, UK

Dada, J.O. and Jagboro, G.O.(2007). *An evaluation of the impact of risk on project cost overrun in the Nigerian construction industry*. Journal of Financial Management of Property and Construction, 12(1), pp.37-44.

Dantata, S. (2007). General Overview of the Nigerian Construction Industry, Cambridge: MSc Thesis for Engineering in Civil and Environmental Engineering: Massachusetts Institute of Technology.

Davies, M.B. (2007). Doing a Successful Research Project; Using Qualitative or Quantitative Methods. New York: Palgrave MacMillan.

Doloi, H.K. (2011). Understanding stakeholders' perspective of cost estimation in project management. *International journal of project managemen*t, 29(5): 622.

Egan, J. (1998). Rethinking construction: the report of the Construction Task Force to the Deputy Prime Minister, John Prescott, on the scope for improving the quality and efficiency of UK construction, London: Department of the Environment, Transport and the Regions Construction Task Force.

Emerging Trends Infrastructure report. (2017). (online). Available at: https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2018/02/emerging-trends-in-infrastructure.pdf. (Accessed 12 April 2017).

Eyiah, A.K. (2004). Regulation and small contractor development: A case study of Ghana. *Centre on Regulation and Competition Working Paper Series. Paper number 80*. Institute for Development Policy and Management, University of Manchester.

Eyiah, A.K. and Cook, P. (2003). Financing Small and Medium-Scale Contractors in Developing Countries: A Ghana Case Study. *Construction Management and Economics, **21**(4)*

Fellows, R. and Liu, A. (2008). *Research methods for construction*. Third ed. Chichester, West Sussex, UK; Malden, MA, USA: Wiley- Blackwell Publication

Ferry, D. J., Brandon, P. S. and Ferry, J. D. (1999). *Cost planning of buildings*. Vol. 7, Oxford, UK: Blackwell Science Ltd.

Flanagan, R. and Norman, G. (1993). *Risk Management and Construction*. Oxford: Blackwell Science Ltd.

Flyvbjerg, B. (2005). Design by deception: The politics of megaproject approval. *Harvard Design Magazine*, **22**, pp.50-9.

Flyvbjerg, B. (2009). Survival of the unfittest: why the worst infrastructure gets built—and what we can do about it. *Oxford Review of Economic Policy*, **25**(3),pp. 344-67.

Flyvbjerg, B. and Stewart, A. (2012). *Olympic Proportions: Cost and Cost Overrun at the Olympics 1960-2012*, Oxford, UK: Saïd Business School, University of Oxford

Flyvbjerg, B., Holm, M. K. S. and Buhl, S. L. (2004). What causes cost overrun in transport infrastructure projects? *Transport Reviews*, **24**(1), pp.3-18.

Flyvbjerg, B., Skamris Holm, M. K. and Buhl, S. L. (2005). How (In)accurate Are Demand Forecasts in Public Works Projects?: The Case of Transportation. *Journal of the American Planning Association*, **71**(2), pp.131-46.

Flyvbjerg, B., Holm, M.S. and Buhl, S. (2006). *Underestimating Costs in Public Works Projects: Error or Lie?* St. Louis, United States, St. Louis: Federal Reserve Bank of St Louis.

Flyvbjerg, B., Holm, M.S. and Buhl, S. (2002). Underestimating costs in public works projects: Error or lie? American planning association. *Journal of the American planning association*, 68(3), pp.279-295.

Flyvbjerg, B., Skamris Holm, M.K. and Søren L. Buhl (2013). Underestimating Costs in Public Works Projects: Error or Lie? St. Louis, United States, St. Louis: Federal Reserve Bank of St Louis.

Fortune, C. and Cox, O. (2005). Current practices in building project contract price forecasting in the UK. *Engineering, construction and architectural management*, 12(5), pp.446-457.

Frimpong, Y. and Oluwoye, J. (2003). Significant Factors Causing Delay and Cost Overruns in Construction of Groundwater Projects in Ghana. *Journal of Construction Research*, 1(2), pp.175 - 187.

Frimpong, Y., Oluwoye, J. and Crawford, L. (2003). "Causes of delay and cost overruns in construction of groundwater projects in a developing countries; Ghana as a case study", *International Journal of Project Management*, Vol. 21 No. 5, pp. 321-326.

Fugar, F. and Agyyakwah - Baah, A. (2010). Delays in Building Construction Projects in Ghana. Australasian *Journal of Construction Economics and Building*, 10(1/2), pp.103 - 116.

Gaetsewe, R., Monyane, T. and Emuze, F. (2015). ppml.url.tw. [Online] Available at:http://www.ppml.url.tw/EPPM/conferences/2015/download/Overruns%20Agai n%20in%20Public%20Projects_Perspective%20from%20Northern%20Cape,%20 South%20Africa.pdf [Accessed 12th November 2017].

Garbharran, H. L., Jeevarathnam, P. & Msani, T., 2013. ajol.info. [Online] Available at: http://www.ajol.info/index.php/actas/article/viewFile/87641/77304 [Accessed 22 June 2016].

Garbharran, H., Govender, J. and Msani, T. (2012). ajol.info. [Online] Available at: http://www.ajol.info/index.php/actas/article/viewFile/87641/77304 [Accessed 19 May 2017].

Ghana Statistical Service Report. (2016). (online). Available at: http://www.statsghana.gov.gh/docfiles/GDP/GDP2016/Revised_Annual_2015_GDP_September%202016%20Edition.pdf. (Accessed 20 Septemper 2016)

Gil, N. and Lundrigan, C. (2012). The leadership and governance of megaprojects. In: *CID Technical Report No. 3/2012*: Centre for Infrastructure Development (CID), Manchester Business School, The University of Manchester, 18.

Graza, J. and Rouhana, K. (1995). “Neural network versus parameter based application in cost.” Cost Eng., 37(2), pp.14–18.

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E. and Tourangeau, R. (2009). Survey Methodology. 2nd Edition. New Jersey: John Wiley and Sons, Inc.

Hagan, M.T. and Menhaj, M.B. (1994). Training feedforward networks with the Marquardt algorithm. IEEE transactions on Neural Networks, 5(6), pp.989-993.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. and Tatham, R.L. (2010). Multivariate data analysis. Upper Saddle River, NJ: Prentice hall.

Harris F. and McCaffer, R. (2006). 6th edn. Modern construction management. Oxford: Blackwell Science.

Hegazy T. (2002). Computer-based construction project management. Upper Saddle River, NJ: Prentice-Hall Inc.

Hinze, J, Selstead, G. and Mahoney, J.P. (1992). Cost overruns on state of Washington construction contracts. *Transportation Research Record* (Issue Number 1351), 87-93.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), pp.55-67.

Huang, G.B., Bai, Z., Kasun, L.L.C. and Vong, C.M. (2015). Local receptive fields based extreme learning machine. IEEE Comput Intell Mag.;10 (2): pp.18–29.

Huang, G.B., Zhu, Q.Y. and Siew, C.K. (2006). Extreme learning machine: theory and applications. Neurocomputing;70. pp:489–501.

Huang, G., Huang, G.B., Song, S. and You, K. (2015). Trends in extreme learning machines: A review. *Neural Networks*, 61, pp.32-48.

Huang, G., Song, S., Gupta, J.N. and Wu, C. (2014). Semi-supervised and unsupervised extreme learning machines. *IEEE transactions on cybernetics*, 44(12), pp.2405-2417.

Huang, G.B. (2014). An insight into extreme learning machines: random neurons, random features and kernels. *Cognitive Computation*, 6(3), pp.376- 390.

Huang, G.B. (2015). What are extreme learning machines? Filling the gap between Frank Rosenblatt's dream and John von Neumann's puzzle. *Cognitive Computation*, 7(3), pp.263-278.

Huang, G.B. and Chen, L. (2007). Convex incremental extreme learning machine. *Neurocomputing*, 70(16), pp.3056-3062.

Huang, G.B. and Chen, L. (2008). Enhanced random search based incremental extreme learning machine. *Neurocomputing*, 71(16), pp.3460-3468.

Huang, G.B., Bai, Z., Kasun, L.L.C. and Vong, C.M. (2015). Local receptive fields based extreme learning machine. *IEEE Computational Intelligence Magazine*, 10(2), pp.18-29.

Huang, G.B., Chen, L. and Siew, C.K. (2006). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans. Neural Networks*, 17(4), pp.879-892.

Huang, G.B., Zhou, H., Ding, X. and Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2), pp.513-529.

Huang, G.B., Zhu, Q.Y. and Siew, C.K. (2004). Extreme learning machine: a new learning scheme of feedforward neural networks. In Neural Networks, 2004. *Proceedings. 2004. IEEE International Joint Conference* Vol. 2, pp. 985-990.

Huang, G.B., Zhu, Q.Y. and Siew, C.K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1), pp.489-501. *IEE Transactions On Engineering Management, Vol. 53, No. 3, August 2006.*

Hughes W., (2001) "Evaluating plans of work", *Engineering, Construction and Architectural Management*, Vol. 8 Issue: 4, pp.272-283

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of decision under risk. Econometrica. 47(2), pp. 263-291

Kaming, P., Olomolaiye, P., Holt, G. and Frank, C. (1997). Factors Influencing Construction Time and Cost Overruns on High rise Projects in Indonesia. *Construction Management and Economics*, 15(1), pp. 83 - 94.

Kasimu, M. A. (2012). Significant factors that cause cost overruns in building construction project in Nigeria. *Interdisciplinary Journal of Contemporary Research in Business*, Vol. 3, No 11.

Kim, G. H., Seo, D. S. and Kang, K. I. (2005). "Hybrid Models of Neural Networks and Genetic Algorithms for Predicting Preliminary Cost Estimates" *ASCE Journal of Computing in Civil Engineering.* 19(2), 208-211

Kolodner, J. (1993). Case-based reasoning. CA, US: Morgan Kaufmann Publishers

Kuo, L. and Mallick, B. (1998). Variable selection for regression models. *The Indian Journal of Statistics, Series B*, pp.65-81.

Lamudi, (2015). lamudi.com.ng. [Online] Available at: http://www.lamudi.com.ng/journal/major - opportunities - in - nigerian - construction - industry/ [Accessed 22 September 2015].

Lapan, S.D. and Quartaroli, M.T. (2009). Research Essentials; An Introduction to Designs and Practices. New Jersey: John Wiley and Sons.

Laryea, S. (2010). 'Contractor project estimates vs. consultant project estimates in Ghana', *Proceedings of the Construction, Building and Real Estate Research Conference of the Royal Institution of Chartered Surveyors*.

Li, K., Peng, J.X. and Irwin, G.W. (2005). A fast nonlinear model identification method. *IEEE Transactions on Automatic Control*, 50(8), pp.1211-1216.

Ling, Y.Y. and Boo, J.H.S. (2001). 'Improving the accuracy estimates of building of approximate projects '. Building Research and Information, vol. 29, no. 4, pp. 312-318.

Lovallo, D. and Kahneman, D. (2003). Delusions of Success: How Optimism Undermines Executives' Decisions. *Harvard Business Review* (July 2003).

Love, P. E. D., Edwards, D. J. and Irani, Z. (2012). Moving beyond optimism bias and strategic misrepresentation: An explanation for social infrastructure project cost overruns. *IEEE Transactions on Engineering Management*, **59**(4), pp.560-71.

Love, P. E. D., Edwards, D. J. and Smith, J .(2005). Contract documentation and the incidence of rework in projects. *Architectural Engineering and Design*

Management, **1**(4), pp.247-59.

Love, P. E. D., Sing, C.P., Wang, X., Irani, Z. and Thwala, D. W. (2012). Overruns in transportation infrastructure projects. *Structure and Infrastructure Engineering*, pp.1-19.

Love, P. E. D., Smith, J., Simpson, I., Regan, M., Sutrisna, M. and Olatunji, O. (2014).Understanding the Landscape of Overruns in Transport Infrastructure Projects. *Environment and Planning B: Planning and Design*.

Mamman, E. and Omazokpia, E. (2014). An Evaluation of Factors Affecting the Performance of Construction Projects in Niger states. *Journal of Environmental Sciences and Resources Management*, 6(1), pp. 34 - 43.

Mansfield, N., Ugwu, O. and Doranl, T. (1994). Causes of Delay and Cost Overruns in Nigerian Construction Projects. International Journal of Project Management, 12(4), pp. 254 - 260.

Mao, L., Zhang, L., Liu, X., Li, C. and Yang, H. (2014). Improved extreme learning machine and its application in image quality assessment. *Mathematical Problems in Engineering, 2014.*

Marzouk M. M. and Ahmed, R. M. (2011). A case-based reasoning approach for estimating the costs of pump station projects, Journal of Advanced Research 2, pp.289-295

Ministry of Finance and Economic Planning Budget Statement Ghana, 2016 (Online) Available at: https://www.mofep.gov.gh/sites/default/files/budget-statements/2016-Budget-Statement.pdf. (Accessed 02 January 2017).

Marzouk, M.M. and Ahmed, R.M. (2011). A case-based reasoning approach for estimating the costs of pump station projects. *Journal of Advanced Research*, 2(4), pp. 289–295.

Monyane, T.G. (2013). encore.tut.ac.za. [Online] Available at: http://encore.tut.ac.za/iii/cpro/DigitalItemViewPage.external?lang=eng&sp=1001373&sp=T &suite=def [Accessed 22nd November 2015].

Moselhi, O., Gong, D. and El-Rayes, K. (1997). "Estimating Weather Impact on the Duration of Construction Activities." *Canadian Journal of Civil Engineering* , 24, pp.359-366.

Mukuka, M., Aigbavboa, C. and Thwala, W. (2014). ppml.url.tw. [Online] Available at: http://www.ppml.url.tw/EPPM/conferences/2014/download/Construction%20Experts%E2%80%99%20Perception%20on%20the%20Causes%20and%20Effects%20of%20Cost%20Overruns%20in%20Johannesburg,%20Gauteng%20Province,%20South%20Africa.pdf [Accessed 10th November 2017].

Mustafa, A. (2010). Research Methodology. Delhi: A.I.T.B.S Publishers.

Navidi, W.C. (2006). *Statistics for Engineers and Scientists*, New York: McGraw-Hill.

Neuman, W.L. (2011). Social Research Methods: Qualitative and Quantitative Approaches. 7th Edition. New Jersey: Pearson Education.

Ng, S. T., Mak, M. M. Y., Skitmore, R. M., Lam, K. C. and Varnam, M. (2001). "The predictive ability of Bromilow's time-cost model." *Construction Management and Economics*, 19, pp.165-173.

Nkado, R. N. (1992). "Construction time information system for the building industry." *Construction Management and Economics*, 10, 489-509.

Nkobane, M. (2012). Causes of Delay and Cost Overruns in Engineering, Procurement and Construction Management Projects in South Africa, Johannesburg: Masters minor dissertations: University of Johannesburg.

Oberlender, G. D. (2000). Project Management for Engineering and Construction. 2nd ed. United States of America: McGraw - Hill Higher Education.

Odeck, J. (2004). Cost overruns in road construction-what are their sizes and determinants? *Transport policy*, 11(1), pp.43-53.

Okmen, O. and Öztas, A. (2010). Construction cost analysis under uncertainty with correlated cost risk analysis model. *Construction Management and Economics*, **28**(2), pp.203-212.

Olapado, A. (2007). A Quantitative Assessment of the Cos t and Time Impact of Variation Orders on Construction Projects. Journal of Engineering, Design and Technology, 5(1), pp. 35 - 48.

Olatunji, A. A. (2010). Influence on Construction Project Delivery Time, Port Elizabeth: PHD Thesis for Nelson Mandela Metropolitan University.

Olawale, Y. and Sun, M. (2010). Cost and Time Control of Construction Projects: Inhibiting factors and Mitigating Measures in Practice. Management and Economics, 28(5), pp. 509 - 526.

Olumide, A. (2009). Sliding Scale Contingencies for the Highway Construction Project Development Process, MSc Thesis for Texas A&M University.

Osland, O. and Strand, A. (2010). The Politics and Institutions of Project Approval-a Critical-Constructive Comment on the Theory of Strategic Misrepresentation. *EJTIR*, **1**(10).

Owolabi, J.D., Amusan, L.M., Oloke, C.O., Olusanya, O., Tunji-Olayeni, P.F., Dele, O., Peter, N.J. and Omuh, I.O. (2014). Causes and effect of delay on project construction delivery time. *International journal of education and research*, *2*(4), pp.197-208.

Oyewobi, L., Ibironke, O., Ganiyu, B. and Ola - Awo, A. (2011). Evaluationg Rework Cost - A Stdudy of Selected Building Projects in Niger State, Nigeria.

Journal of Geography and Regional Planning, 4(3), pp. 147 - 151.

Peter F. (2005). *Construction Project management, An Integrated Approach.*, 1^{ST} edn., Taylor & Francis, London and New York.

Raisbeck, P., Millie, R. and Maher, A. (2010). Assessing integrated project delivery: a comparative analysis of IPD and alliance contracting procurement routes. In:Egbu, C. (Ed) Procs 26th Annual ARCOM Conference,6-8 September 2010, Leeds, UK, Association of Researchers in Construction Management, pp.1019-1028.

Ramabodu, M. S. and Vester, J. (2013). Factors that Influenced Cost Overruns in South African Public Sector Mega projects. *International Journal of Project Management,* 5(12), pp. 48 - 56.

Rao, C.R. and Mitra, S.K. (1971). Generalized inverse of matrices and its applications.

Rice, J. (2006). *Mathematical statistics and data analysis*, Nelson Education.

Ross, A. and Williams, P. (2013). *Financial Management in Construction Contracting*. Sussex, UK: Wiley-Blackwell.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), p.533.

Cheung, S.O. and Yiu, T.W. (2006). Are Construction Disputes Inevitable? *IEEE Transactions on Engineering Management*, *53*(3), pp.456-470.

Seeley, I.H. (1983). 4^{th} edn. Building Economics. London: Macmillan.

Shenshar, A. J. and Dvir, D. (2007). Reinventing Project Management:The Diamond Approach to Successful Growth and Innovation. Harvard Business School Press.

Skitmore, M. and Marston, V. (1999). Cost Modelling, Taylor & Francis, E & FN Spon 11 New Fetter Lane, London EC4P 4EE

Skitmore, M. and Wilcock, J. (1994). Estimating processes of smaller builders. *Construction management and economics*, 12(2): 139.

Skitmore, R. M. and Ng, S. T. (2001). "Australian project time-cost analysis: statistical analysis of intertemporal trends." *Construction Management and Economics*, 19, pp.455-458.

Skitmore, R. M. and Patchell, B. R. T. (1990). "Developments in contract price forecasting and bidding techniques." *Quantity surveying techniques: New directions*, P. S. Brandon, ed., BSP Professional Books, Oxford, U.K.

Smith, N. J. (1999). Managing risk in construction projects Oxford: Blackwell Science.

Smith, N. J. (2002). 2nd edn. Engineering Project Management. Oxford: Blackwell Science.

Tabachnick, B. G. and Fidell, L. S. (1996). Using Multivariate Statistics. 3rd edition. New York: HarperCollins.

Tam, C.M. and Fang, C, F. (1999). Comparative cost analysis of using high performance concrete in tall building construction by artificial neural networks. ACI Structural Journal;96 (6): pp.927–36.

Tang, J., Deng, C. and Huang, G.B. (2016). Extreme learning machine for multilayer perceptron. *IEEE transactions on neural networks and learning systems*, 27(4), pp.809-821.

Tibshirani, R., James, G., Witten, D. and Hastie, T. (2013). *An introduction to statistical learning with applications in R.*

Ubani, E. and Ononuju, C. (2013). A Study of Failure and Abondonment of Public Sector - Driven Civil Engineering projects in Nigeria: An Emperial Review. *American Journal of Science and Industrial Research,* 14(1), pp. 75 - 82.

Wachs, M. (1989). When planners lie with numbers. *Journal of the American Planning Association*, **55**(4), pp.476.

Wachs, M. (1990). Ethics and advocacy in forecasting for public policy. *Business and Professional Ethics Journal*, **9**(1-2), pp.141–157.

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

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of personality and social psychology*, **39**(5), pp.806.

Westring, G. (1997). “Ghana public procurement reform, an audit report prepared for the World Bank.” Advokatfirman Cederquist KB, Stockholm.

Wilamowski, B.M. and Yu, H. (2010). Neural network learning without back-propagation. *IEEE Transactions on Neural Networks*, 21(11), pp.1793-1803.

Xiao, H. and Proverbs, D. (2002). “The performance of contractors in Japan, the UK and the USA: a comparative evaluation of construction cost”, *Construction Management & Economics,* Vol. 20 No. 5, pp. 425-435.

Yao, X. (1993). A review of evolutionary artificial neural networks. *International journal of intelligent systems*, 8(4), pp.539-567.

Zhang, P., Song, D., Wang, J. and Hou, Y. (2014). Bias–variance analysis in estimating true query model for information retrieval. *Information Processing & Management*, 50(1), pp.199-217.

Zong, W., Huang, G.B. and Chen, Y. (2013). Weighted extreme learning machine for imbalance learning. *Neurocomputing*, 101, pp.229-242.