ARAB ACADEMY FOR SCIENCE, TECHNOLOGY
AND MARITIME TRANSPORT
College of Engineering and Technology

Parametric Cost Estimating of Sterile Building Using Artificial Neural Network & Genetic Algorithm Model

Thesis Submitted in Partial Fulfillment of the Requirements of the Degree of

"Master of Construction & Building Engineering"

Submitted by

Eng. Mohamed SaadAllah Zahran
B.Sc. Civil Engineering
Faculty of Engineering
Ain-Shams University
Cairo, Egypt

Supervised by

Prof. Abdelmonem Yassin Sanad
Professor, Building and Construction Dept., Faculty of Engineering, Arab Academy for Science and Technology and Maritime Transport, Cairo, Egypt.

Prof. Hossam EL-Deen H. Mohammed
Professor, Construction Engineering Dept., Faculty of Engineering, Zagazig University
Zagazig, Egypt.

April 2015

ARAB ACADEMY FOR SCIENCE, TECHNOLOGY

AND MARITIME TRANSPORT

College of Engineering and Technology

Parametric Cost Estimating of Sterile Building Using Artificial Neural Network & Genetic Algorithm Model

Submitted by

Eng. Mohamed SaadAllah Zahran
B.Sc. Civil Engineering
Faculty of Engineering, Ain-Shams University
Cairo, Egypt

A Thesis
Submitted in Partial Fulfillment of the Requirements
of the Master's Degree in

Construction and Building Engineering

Supervised by

Prof. Abdelmonem Yassin Sanad **Prof. Hossam EL-Deen H. Mohammed**

______________________ ______________________

Examined by

______________________ ______________________

April 2015

قال تعالى في كتابه الكريم

وقل رب زدني علما

١٤٣٥

طـــه – آيه 114

DEDICATED TO MY BELOVED

PARENTS, BROTHER, SISTERS, WIFE, KIDS AND FRIENDS

Who are the source of my inspiration, encouragement, guidance and happiness, and who share my goals and aspirations May Almighty ALLAH Bless and Protect them.

A special feeling of gratitude to my beloved parents, who dead during this research, my words can’t demonstrate how much I love you. GOD Bless You.

ACKNOWLEDGMENT

First and Foremost, thanks to ALLAH, God of all, for everything happened in my life. All the praises and thanks are to Almighty ALLAH (SWT), the most gracious, the most merciful, who gave me the knowledge, courage and patience to accomplish this research. May the peace and blessings of Allah be upon Prophet Muhammad (PBUH).

It is my great pleasure and honor to have “Prof. Hossam El Deen H. Mohammad” as my thesis supervisor, who supported and guided me throughout my research and offered his precious time, experience and advice.

Moreover, I would like to thank “Prof. Abdelmonem Yassin Sanad” Building and Construction Dept., Faculty of Engineering, AASTMT for his care and valuable support during my research.

Special thanks go to “Eng. Hisham Howeedy from Howeedy Consultant” who supported me with the real life experience in the Construction Management field, and provided me with all data required during my research.

ABSTRACT

Conceptual cost estimating is one of the most important and challenging activities during project planning, which occurs at the early stages of a project life where limited information is available and many unknown factors affecting the project costs. While many studies have indicated the importance of accurate conceptual cost estimates, there has been little effort directed at improving the conceptual cost estimate processes, especially for construction projects in the industrial field.

The main objective of this thesis is to develop an accurate and reliable parametric cost estimating model, this model can be used by organizations and individuals involved in the planning and execution of industrial construction projects. The proposed model will be limited to project encompassing special hygienic buildings called sterile buildings, such as pharmaceutical, food and dairy industrial projects.

To build that model, the most important factors affecting cost in construction projects were identified in this thesis based on a comprehensive survey among a collected sample of construction experts only. These factors, were found to be as follows, (1) Currency exchange rate during the month of studying the project. (2) Egypt Consumer price index during the month of studying the project, (3) Accumulative built-up area (total area), (4) Accumulative Sterile areas (total area), (5) Other supplementary buildings (water tank, administration, warehouse, … etc) (total area), (6) Project location, (7) Target market (regional or international), (8) International insurances (if any), (9) Desired level of contractor's prequalification, (10) Desired Duration for the project, (11) Desired structural system, (12) Buildings closeness (attached, semi-attached or separated), (13) Project Status (renovation, extension or …. etc).

Thereafter, an Artificial Neural Network model was developed in order to help in cost estimating process. The previously identified thirteen factors that were found to be the most important cost factors represent the main inputs of the

proposed model. The output of this model is the expected construction cost of the project. Moreover, the average percent of Error (%Error) can be used as a cost contingency.

The model has been developed and trained based on the collected actual data of a collected sample of similar industrial projects, eighteen projects. The main steps of the model development were deeply discussed. Moreover, the validity of the proposed model was tested based on a sample of four projects that haven't seen by the model before. Furthermore, the model was validated using two entirely new projects. The results of such validation clearly provides a good indicator regarding the ability of the proposed model to predict the construction cost of any future industrial construction project at an appreciated degree of accuracy.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENT v

ABSTRACT vi

LIST OF FIGURES x

LIST OF TABLES xi

CHAPTER 1: INTRODUCTION 2

1.1 Overview ... 2
1.2 Research Motivation ... 2
1.3 Problem Statement ... 5
1.4 Research Objectives ... 6
1.5 Research Scope ... 7
1.6 Research Limitations ... 7
1.7 Thesis Organization ... 8

CHAPTER 2:LITERATURE REVIEW ... 10

2.1 Introduction ... 10
2.2 Definition of Cost Estimation ... 11
2.3 Types of Cost Estimation ... 11
2.3.1 Conceptual Cost Estimate ... 12
2.3.2 Detailed Estimate ... 13
2.3.3 Other Types of Estimate ... 14
2.4 Conceptual Cost Estimating Methodologies ... 15
2.4.1 Analogous Method ... 16
2.4.2 Assembly Method ... 17
2.4.3 Unit Method ... 18
2.4.4 Parametric Method ... 18
2.5 Procedures of Parametric Estimating ... 27
2.6 Cost Drivers for Parametric Estimating of Construction Projects ... 30
2.7 Conclusion ... 32

CHAPTER 3: ARTIFICIAL NEURAL NETWORK ... 34

3.1 Introduction ... 34
3.2 Concept ... 34
3.3 Structure ... 35
3.4 Genetic Algorithms ... 38
3.5 Problems with Neural Networks ... 39
3.6 Applications in Cost Estimation ... 40
3.7 Conclusion ... 43

CHAPTER 4: DATA COLLECTION AND ANALYSIS 45

4.1 Introduction 45

4.2 Objective 45

4.3 Methodology 46

4.4 Questionnaire 46

4.4.1 Definition 46

4.4.2 Questionnaire Design 47

4.4.3 Questionnaire Sample Size 55

4.4.4 Questionnaire Distribution and Collection 57

4.4.5 Questionnaire Response 57

4.4.6 Questionnaire Analysis 60

4.5 Historical Data Collection 75

4.6 Conclusion 78

CHAPTER 5: DEVELOPMENT OF ARTIFICIAL NEURAL NETWORK MODE 80

5.1 Introduction 80

5.2 Design of ANN Model 80

5.2.1 Selection of Used Software(s) 80

5.2.2 Neural Network Architecture 81

5.2.3 Neural Network Type 83

5.3 Modeling Data 84

5.3.1 Data Categorization 84

5.3.2 Model Execution 84

5.4 Model Implementation 90

5.5 Trial and Error Practices 91

5.6 Model Testing 93

5.7 Model Validation 94

5.8 Model Limitation 95

5.9 Conclusion 95

CHAPTER 6: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS . 97

6.1 Introduction 97

6.2 Summary 97

6.3 Conclusion 98

6.4 Recommendations for Future Studies 99

REFERENCES 102

APPENDIXES 106

APPENDIXES 107

APPENDIX (A) 107

APPENDIX (B) 113

ملخص البحث 124

LIST OF FIGURES

Page

Figure (1.1): The Freiman Curve [Adapted by Phaobunjong 2002]....................... 4
Figure (2.1): Estimate Types Throughout a Project's Life-Cycle (Conceived From Phaobunjong 2002 15
Figure (2.2): Parametric Cost Estimating Procedure.. 28
Figure (3.1): Artificial Neural Network Structure.. 35
Figure (4.1): Categories Percentage 51
Figure (4.2): Classification of Respondents Based on Their Experience............. 58
Figure (4.3): Pareto's Principle "The 80-20 Rule" 70
Figure (5.1): Input Data 85
Figure (5.2): Scaled Data... 85
Figure (5.3): Weighted Input Summation 86
Figure (5.4): Activated Summation.. 87
Figure (5.5): Activated Output Summation.. 88
Figure (5.6): Predicted Cost.. 89
Figure (5.7): Error Calculation 90
Figure (5.8): The Assigned Cells for Learning Role and Monitoring................... 90
Figure (5.9): RMSE for Each Trial 93

LIST OF TABLES

Page

Table (2.1): **Related Studies and Their Relevant Cost Drivers [Adapted from Stoy et al. (2008)]....31**

Table (3.1): **Mathematical Activation Functions Used in ANNs Artificial....37**

Table (4.1): **Cost Factors Collected From Literature Review....47**

Table (4.2): **Cost Factors Derived From Interviews With Experts....48**

Table (4.3): **Categorization of Factors....50**

Table (4.4): **Qualitative Scale Used in Questionnaire....52**

Table (4.5): **Quantitative Scale Used to Interpret Questionnaire....52**

Table (4.6): **The Questionnaire Form....53**

Table (4.7): **Table of Responses....58**

Table (4.8): **Importance Index of Cost Factors....62**

Table (4.9): **Ranking of Cost Factors....65**

Table (4.10): **The Calculations Leading to the Highest Important Factors....67**

Table (4.11): **The Highest (Most) Important Factors....71**

Table (4.12): **The Cost Indicators....74**

Table (4.13): **The Data Collection Form....75**

Table (4.14): **The List of Accepted Projects....76**

Table (4.15): **The New Projects....77**

Table (4.16): **The Assigned Input Characters....78**

Table (5.1): **The Input Layer and the Determination Value for Each Cost Indicator....82**

Table (5.2): **Summary of Trials and Error Practices....91**

Table (5.3): **Detailed Log for Trials and Error Practices....92**

Table (5.4) : **Testing Results of Developed Neural Networks Model....94**

Table (5.5): **Validation Results of Developed Neural Networks Model....95**

CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 Overview

Cost is an important aspect to everyone, especially to those in the construction industry. Every feasibility study, for any investment (project), requires accurate cost estimation in order to inspire the decision makers to either move forward or cancel this investment. Moreover, cost estimating is one of the most important tools for construction project management. For instance, it provides the founders with an excellent image for the predicted cash flow throughout the project life cycle. The reliability of such estimation is greatly affected by many uncertain but predictable factors. So, it is worthy to pay extensive effort to achieve a more reliable estimation of the final construction cost.

1.2 Research Motivation

Conceptual cost estimating is one of the most important activities during project planning. Every project begins its life from concepts proposed by the owner and refined by the designers. Planning decisions in this early stage of any project are vital, as it can have the biggest influence on the subsequent outcome of the project. Planning decisions are based on several planning activities, one of which is the conceptual cost estimating. Conceptual cost estimating is the determination of the project's total cost based only on general early concepts of the project. Like all other planning activities, conceptual cost estimating is a challenging task. This is due to the nature of planning, which occurs at the early stages of a project where limited information is available and many unknown factors affecting the project costs. Moreover, the uncertainties plaguing the construction industry further complicate the planning processes.

While studies have indicated the importance of accurate conceptual cost estimates, there has been little effort directed at improving it, especially for construction projects in the industrial field. Estimating construction costs of industrial projects can be difficult as most projects are unique. Any industrial construction project is a very complex undertaking, which can be composed of hundreds or thousands of construction work items. These work items are often performed by workers or crews from different crafts, utilizing various materials of many different varieties. Due to these complexities, numerous factors can affect the construction processes and ultimately their costs.

The complexity of the industrial construction project and the lack of time and information allocated for conceptual cost estimating often lead to a poor performance of the estimate. The outcome of an estimate can be accurate, underestimate, or overestimate. An accurate estimate generally results in the most economical project cost, while an underestimate and overestimate often lead to greater actual project expenditures (Phaobunjong 2002). This concept can be seen in figure (1.1) the Freiman curve. Underestimate means that the design and specifications cost more than they are estimated. It is also often a result of poor planning and estimating, whereby substantial cost items may be omitted. This unrealistic estimate leads to project delays, terminations, reorganization, and re-planning, which usually results in significant cost growth. On the other hand, an overestimate can be as bad as an underestimate. Although the project will be feasible due to more than adequate funding, the allocation of extra budget will often be completely spent. In this way, the project may seem to finish under the budget, but in truth may cost more than it has to. In this manner, only an accurate, realistic estimate can lead to achievable cost or a truly successful project cost performance (Phaobunjong 2002).

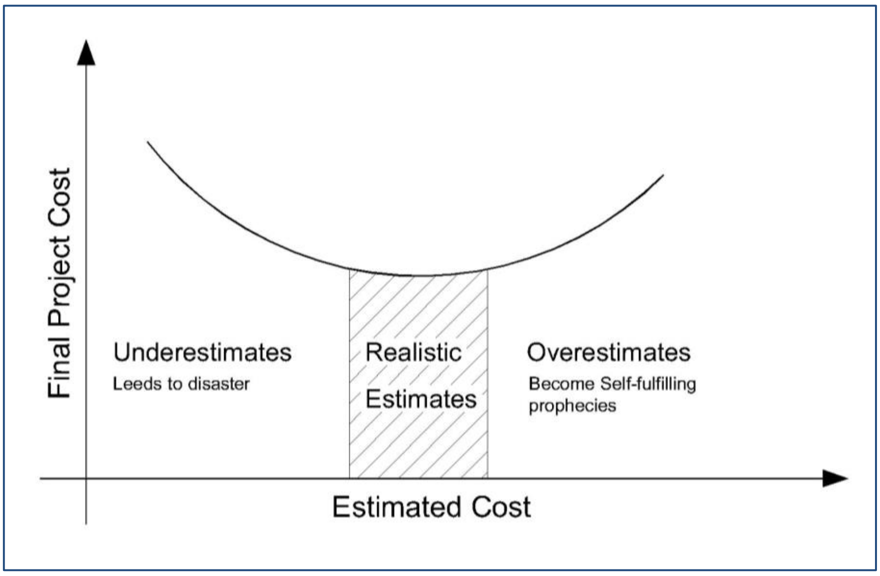


Figure (1.1): The Freiman Curve [Adapted by Phaobunjong 2002]

Current practices in the conceptual cost estimating for industrial projects can range from an educated guess by an experienced estimator to a systematic complex cost estimating model. How an estimate is done is determined largely by the time and effort that is provided to carry out the estimate along with the available resources. An educated guess by an experienced estimator may be the fastest method and requires the least amount of resources, but it can also be the most subjective and unreliable. On the other hand, a systematic complex cost estimating model, although more accurate and reliable, requires a lot of resources for its development and implementation, so that only large organizations can afford.

In addition, construction cost estimating is generally an organization problem rather than an industry problem. That is, the construction processes can be unique to an organization, such that future work performances undertaken by the organization are more related to the organization's similar past experiences than to the experiences of another organization doing similar work or the industry as a whole. In this way, a cost estimating model that works well for one organization may not necessary work for another. Therefore, to improve estimating performances, an organization must develop its own cost estimating model, such that its past experiences can be captured and utilized to predict future performances (Phaobunjong 2002).

The limited research in the area of conceptual cost estimating, especially for the industrial construction projects, and the need for a better conceptual cost estimating methodology and tools are the main incentives for initiating this study.

1.3 Problem Statement

Previous argument highlighted the need for estimating methodologies that utilize relevant historical projects data for the development of accurate estimates. Current practices of conceptual cost estimating for the industrial projects have been performed mainly through the experience of the estimator. One of the main problems hindering this research effort is the lack of extensive historical data on previously completed projects. Limited data have been collected and recorded about industrial construction projects. This is primarily because of the limited knowledge on the importance and potential applications of these historical data.

In addition, there is also a lack of knowledge on what specific data to collect, how to collect them, and what can be done with them. Due to the limited research in this area, there is no systematic method of data collection. Without systematic definition and clarification, the recorded information can be of limited use and is often inconsistent from project to another.

Moreover, Conceptual cost estimating in the industrial construction project is a difficult and generally a subjective process with no standard methodology of practice. There is a need to develop a systematic methodology for conceptual cost estimating, so as to standardize and facilitate the estimating process, making the approach more objective. In this way, the quality of the estimate produced can be more accurate and consistently enhancing.

Besides, the absence of researches directed to industrial construction projects in field of pharmaceutical and food industries, that means there are no related studies to enhance or to be as a guide line for the newest researches. This condition exaggerates the problems associated with cost estimating process for this type of projects.

1.4 Research Objectives

The main objective of this thesis is to develop an accurate, reliable and practical model for systematic parametric cost estimating. This model can be used by organizations and individuals involved in the planning and execution of a certain type of industrial buildings projects.

In order to build such model, the cost factors affecting projects' construction cost that must be firstly investigated. However, there some factors are common in all buildings, there are several factors are altering from type of buildings to another, these factors characterize the building type and use.

The intended modeling methodology will be based on using Artificial Neural Network (ANN) technique to develop the required cost estimating model. Besides, genetic algorithms based software will be used to search for near-optimal solution of the network. The model will use Microsoft Excel spread sheets as a data base information modeling and Evolver software as genetic algorithm based program.

Accordingly, the main objective of this thesis can be concluded in the following items:

1. Identifying the common cost factors affecting the construction cost of industrial projects in the Egypt.
2. Determine the special cost factors affecting the construction cost of industrial projects which comprising sterile buildings, such as pharmaceutical and food industrial projects.
3. Ranking these factors according to their relative importance to find out the most important factors.
4. Selecting cost indicators which can represent the most important cost factors.
5. Collecting of historical data for construction of industrial projects in Egypt.

6. Developing of Artificial Neural Network model that can be used for parametric cost estimating of industrial construction projects, these projects have a special character called sterile buildings, which means buildings with no microbial load as pharmaceutical, food and dairy industrial projects.

1.5 Research Scope

The scope of this thesis was directed to industrial construction projects, specifically aimed at project encompassing special hygienic buildings called sterile buildings such as pharmaceutical, food and dairy industrial projects.

In addition, the design and construction of those types of buildings require additional considerations to comply with the rules and regulations for both regional and international authorities and market. The difficulties associated with the adaptation of these buildings according to the authorities' regulations strongly affect the actual construction cost.

1.6 Research Limitations

The scope of this thesis is limited to the development of a parametric cost estimating model for conceptual cost estimating of industrial construction projects, such as pharmaceutical, food and dairy projects. These projects have special facilities called sterile buildings.

This thesis is limited to the investigation and analysis of previous projects that are similar and pertained to the main focus of the thesis; these projects were constructed in Egypt within the period from year 1999 to 2013 in the industrial zones.

Moreover, the research is concerned with the total construction costs of the project excluding the analysis of other costs associated with fixed and movable equipment, engineering work, and construction contingencies. In addition, these associated costs are naturally inherited in the total cost of the project.

1.7 Thesis Organization

This dissertation is organized into six chapters. It also includes two appendices containing supporting information used in the study, and the developed computer model. Chapter one presents the thesis introduction. Chapter two provides a summary of a very comprehensive literature review that focuses on conceptual cost estimating. Chapter three provides a discussion about the artificial neural network (ANN) and the genetic algorithms (GA) as a solving technique for the ANN. Chapter four presents a brief overview of the data collected from experts for this thesis and the analysis performed on the data to assist this effort, the conducted analysis methodologies and steps are presented along with the results, concurrently provides the available data collected from previous projects. Chapter five focuses on the development of the parametric cost estimating model through the use of artificial neural network and genetic algorithms, the validation of the model is also presented along with the conclusions and limitations. A summary, conclusions, and recommendations for additional research are discussed in Chapter six. Appendix (A) encloses samples of the questionnaire responses, while appendix (B) includes the collected historical data from similar projects constructed in the past.

.

CHAPTER 2

LITERATURE REVIEW

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Cost estimation is probably the most crucial function to the success of construction projects. At all project phases, different types of cost estimating from conceptual to detailed were conducted for different purposes according to the project phase regarding the data availability at the phase.

This chapter firstly illustrates the different types of cost estimating and describes main characteristics to recognize the attributes and use of each type such as; appropriate project phase to use these techniques, amount of available information, associated difficulty level and expected accuracy of the estimate type.

Thereafter the parametric cost estimating will be extensively elaborated to determine the advantages and drawbacks inherent in such estimating techniques that used in case of either lake or absence of available information of the project at the early stage, almost at conceptual phase of construction projects. Besides, this chapter will represent the techniques used to conduct the parametric cost estimating. Afterward illustrate the procedure of using parametric cost estimating and the main parameters that affecting the use of such type of cost estimating.

Finally, this chapter surveys the previous researches that developed in scope of project cost estimation using not only old techniques, but also new techniques such as; computer based techniques and techniques that use artificial intelligence bases. At the last this chapter concludes the main cost factors that affect parametric cost estimating process of construction projects, these factors called cost drivers.

2.2 Definition of Cost Estimation

Cost estimating involves developing an approximation (estimate) of the costs of the resources needed to complete project activities (PMBOK 2008). A cost estimate can also be defined as an evaluation of all costs of the elements of a project or effort as defined by an agreed upon scope. It is also an assessment based on specific facts and assumptions of the final cost of a project. Throughout the project life cycle, as the project progresses, it goes through the different phases of development. Several different types of estimates are required as a project is conceived, designed, engineered, and constructed. A detailed estimate based on computed material quantities, or quantity take-off, cannot be made at the conceptual stage or preliminary design stage, because the plans and specifications have not been developed. At these phases, the project has not been fully defined. Alternative methodologies are required to estimate the cost of the project during these early phases. To manage the project cost, estimates are performed at the different stages along the project phases in order to monitor and control the cost of the project, and to make timely important project decisions, such as those relating to project feasibility. The objective of the cost estimate is to calculate and predict the most probable cost of the project based on the available information at the time the estimate is performed (Arafa 2011).

2.3 Types of Cost Estimation

There are many types of cost estimates that can be conceived on a project, considering the project phase as well as the purpose of the estimate, each type having different levels of accuracy. The estimating process becomes increasingly more expensive as more detailed and accurate techniques are applied. Analysis of different classifications of estimates concludes that there are two main types of estimates, the conceptual and the detailed estimate.

The selection of a certain type or method of cost estimating depends mainly upon the purpose of the estimate. However, the available data and information, at time of forecasting the cost estimate, are also constraining the selection of cost estimate type. The pre-mentioned types of estimate will be described regarding the project phases in the following items.

2.3.1 Conceptual Cost Estimate

Conceptual estimate is also known as a top-down, order of magnitude, feasibility, quickie, analogous, or preliminary estimate. It is the first serious effort made at attempting to predict the cost of the project. A conceptual estimate is usually performed as part of the project feasibility analysis at the beginning of the project. In this way, the estimate is made with limited information on project scope, and is usually made without detailed design and engineering data. The accuracy range is expected to be +50% to –30%.

The conceptual estimate, pre-design estimates are usually performed with limited or no design and engineering information. Conceptual estimates are frequently prepared when design and engineering have not even started. Project information available in these early stages is usually high-level information, such as number of building occupants, gross square footage area, or building enclosed volume. The estimator may have to make such an estimate from rough design sketches, without dimensions or details and from an outline specification and schedule of the owner's space requirements. The estimate derivation methodologies are usually those relying on historical information to predict future cost of the new project, such as referencing to previously completed projects to estimate the cost of a new project that is similar in nature (Arafa 2011).

This type of estimate is conducted at the early stage where project budgets are to be decided. The preliminary cost estimate can serve several purposes, including: 1) it supplements or serves as the owner's feasibility estimate; 2) it aids the Architect/Engineer in designing to a specific budget; and 3) it assists in the establishment of the owner's funding. The most common methodologies of this type are; analogous, assembly and parametric, these methodologies will be elaborated extensively in section (2.4).

2.3.2 Detailed Estimate

A detailed estimate is also known as a bottom-up, fair-cost, or bid estimate. A detailed estimate is performed after the completion of project design and specifications, which clearly indicates the required quality of materials and workmanship. A detailed estimate is prepared from a well-defined design and engineering data. A fair-cost estimate is carried out by owner for bid evaluations, contract changes, extra work, legal claims, permits and government approvals. A bid estimate is prepared by contractor to be submitted to the owner as the proposed cost of carrying out the construction work. The expected accuracy for a definitive estimate is within a range of +15% to –5%.

A detailed estimate is a post-design estimate. Unlike the pre-design category, the post-design cost estimate relies on completed design and engineering data. It is usually performed when project design and engineering have been completed. The estimating methodology is more complex and detailed, requiring careful tabulation of all material quantities required for the project as well as the identification of all cost items. These quantities are then multiplied by selected or developed unit costs, and the resulting sum represents the total estimated construction cost of the project (Emsley 2007).

This type of estimate is conducted at an intermediate project stage by dividing a project into convenient functional elements (excavation, foundation work, concreting, …etc.) and individually pricing each of these elements. Element breakdown can serve several purposes: 1) to reveal the distribution of costs of the constituent elements of the project; 2) to relate the cost of a constituent element to its importance as a part of the whole project; 3) to compare the cost of the same element in different projects; and 4) to enable a determination of how costs could have been allocated to obtain a better project.

2.3.3 Other Types of Estimate

Between conceptual estimate and detailed estimate, other estimates are performed as the project becomes more defined and more information becomes available. Those estimates are required to assess the most accurate expected cost of the project at the time the estimate is carried out. They may be referred to as budget, appropriation, control, semi-detailed, design or engineering estimates, and are carried out for the purpose of assigning project budgets, and to monitor and control project costs.

In addition to the above listed pre-construction estimates, other estimates are also performed during the project's construction phase or after the construction completion to assess the final actual cost of the project. The estimates in this stage are known as definitive estimates. These estimates are updates of the detailed estimates with emphasis in actual rather than projected construction cost (Lekan 2011).

The various estimates discussed above are carried out in sequence, as the previous cost estimate being an input to the next one. The estimates are successively refined incorporating new information, thus keeping the continuously updated estimate that becomes the budget, available for the control processes. As the project progresses, the amount of unknowns and uncertainties decreases, while the level of details and the project information increases. In this way, the accuracy of the estimate improves as it moves from conceptual to detailed estimate.

Figure (2.1) illustrates different types of cost estimate along with their appropriate project phases, associated difficulty levels, and the expected cost estimate accuracy.

Estimate Type	Conceptual	Other	Detailed
Project Phase	Concept		Construction
Available Information	Limited →		Detailed
Difficulty Level	High →		Low
Expected Accuracy Range	+50% -30%		+15% -5%

Figure (2.1): Estimate Types Throughout a Project's Life-Cycle (Conceived From Phaobunjong 2002

According to the previous discussion we can realize that, at all estimating levels, the embedded difficulties often result in many mistakes and errors in judgment. However, as shown in figure (2.1), conceptual estimating exhibits the lowest accuracy level due to the lack of project information and the high level of uncertainty at this early stage of a project. In addition, the political, site, environmental, and technological risks that are extremely difficult to predict in this stage of project life cycle.

With the above discussions on the various estimating types in construction projects, the next sections focus on conceptual cost estimating, presenting an in-depth discussion on its methodologies.

2.4 Conceptual Cost Estimating Methodologies

The vital element of conceptual cost estimating input is the estimating methodology. The estimating methodology is required to evaluate the project scope and manage and analyze the gathered information to produce the conceptual cost estimate. Conceptual cost estimating methodology can vary considerably from one type of construction to another. For example, an office building construction project will be estimated differently than an industrial process plant construction project. Obviously, the nature of both projects is different, but more importantly, the cost components associated with the two projects are different, consequently, each of the two construction projects will require two different methodology approaches for estimating the project construction costs. Conceptual cost estimating spans across three methodologies, those methodologies encompass all cost estimates made in the early phases of a project. Whereas, the selection of the appropriate estimate methodology, relies on the level of information available when the estimate is performed. Consequently, the methodologies used for conceptual cost estimating can range from a simple process requiring the least amount of information to a complex process where more information is needed or is explicitly assumed. Generally the most sophisticated and accurate procedures have been developed for large projects in the industrial construction sector by the owner or design-construct firms. Most of the existing conceptual cost estimating methods for building construction project falling into one of the following categories (Charles 2006 and PMBOK 2008):

1. Analogous Method
2. Unit Method
3. Parametric Method
4. Assembly Method

The four methods are presented above in the order of increasing methodology complexity, information required, effort required and estimate accuracy. The four methods of conceptual cost estimating also represent the continuum of conceptual cost estimating methodologies, whereby the methodologies are subsequently refined from a more qualitative method to a more quantitative method.

2.4.1 Analogous Method

Analogous estimating is a form of expert judgment. It is a qualitative method and is very subjective. It involves using the actual cost of a previous similar project as the basis for estimating the cost of the current project (PMBOK 2008). One or several projects may be used as reference projects. There is generally very little quantitative basis in the project selection process, which is often limited to the estimator’s judgment and experiences. The method usually involves very little calculation. Any calculation performed is typically limited to using indexes to adjust the costs and determining the average cost for similar projects. Due to its simplistic nature and limited project scope information required, the estimate has poor accuracy but serves its purpose for preliminary economic feasibility consideration. However, analogous estimating is generally the least demanding and least costly method. Analogous estimating is most reliable when the reference project or projects are really similar in fact and not just in appearance, and that the estimators involved in preparing the estimates have the needed expertise (Charles 2006 and PMBOK 2008).

2.4.2 Assembly Method

The assembly method is the most complex and demanding methodology for conceptual cost estimating. Assemblies are smaller collections of related items and tasks that have been combined to form a distinct function or task. For example, an assembly for a concrete floor slab on-grade may include the quantities or costs required for formwork, reinforcing steel, and concrete per square foot of the slab area. The assemblies, when arranged into meaningful building systems, form the building blocks for the project. In this way, the development of the assembly model can be as complicated as the task of performing the detailed estimate. Assembly method can be an accurate method of estimating as it is a very detailed method. However, developing an assembly model generally is a major undertaking. The primary input to the model development is the detailed records of completed projects. Analyzing and maintaining the detailed records in the database demands a tremendous effort and resources such that generally a small company cannot afford. Due to the amount of effort required for developing the model, the use of the assembly method is typically restricted to large organizations involved in repetitive, specialized construction projects (Charles 2006).

2.4.3 Unit Method

The refinement for the analogous method is the unit method and parametric method. These methods of estimating are characterized by the fact that project cost is related to one or several project units or parameters. Units used in building construction project are readily quantifiable building characteristics, such as number of parking spaces for automobiles in a parking facility, or number of beds in a dormitory or hospital type facilities. However, the most common unit cost estimate for a building is the cost per square foot. Area is perceived to have a powerful effect upon costs, and thus its popularity. The unit method of cost per square foot is the most popular of all estimating methods. Unlike the parametric method, the unit method generally refers to the estimating method whereby the estimates are determined from predetermined unit costs (PMBOK 2008).

2.4.4 Parametric Method

Parametric cost estimation is a way of estimating costs using Cost Estimating Relationships (CERs), along with mathematical algorithms and other logic to establish a cost estimate. Moreover, the parametric method seeks to evaluate the costs of a project from the parameters characterizing the project but without describing it completely. Simply put, if the original cost is known for similar project of a proposed project, users can take simple inputs and generate a model that will produce a cost that is reasonably accurate. Ease of use, once the model is constructed, is the key advantage of parametric cost estimation. Parametric methods use the knowledge of a certain number of physical characteristics or parameters such as the type of the building, area, the number of repeated floors and the number of inputs or outputs. This allows an estimate to be created without having the knowledge of detailed information. Parametric estimates are based on historical data and mathematical expressions relating cost as the dependent variable to selected, independent, cost-driving variables through regression analysis.

The implicit assumption of this approach is that the same forces that affected cost in the past will affect cost in the future. Based on the information given above, the main benefit of parametric cost estimation is its ease of use. Once the model has been defined, application of the method is straightforward. Data is easy to implement, and any user can look at a conceptual assumptions and find the input data needed. By taking general broad attributes, a fairly accurate estimate can be generated. This allows an estimate to be made without the presence of a detailed conceptual design. Some disadvantages of parametric cost estimation are that the model itself can be very difficult to develop. To begin, one needs to first develop appropriate parameters, and this person usually has to be an expert on the topic. Next, historical data must be found, as well as relationships that can take vague inputs and produce a cost that is as accurate as possible. On top of being difficult to develop, accuracy is another great drawback of this technique. Other methods, while more complicated for the user, are generally more accurate.

The term "Parametric Estimating", however, may mean different things to different researchers. For instance, it may mean the determination of the life cycle cost of a system from a mathematical model containing a number of parameters and based on case histories of similar projects. On the other hand, it could mean the cost estimation of any system, made up of aggregated components (PMBOK 2008).

Parametric cost estimate, however its drawbacks and difficulties, plays an important role for all stakeholders of the project; Parametric estimates are used to assist: 1) owners in making go-no-go decisions, 2) owners and estimators in the definition of project scope and characteristics, 3) owners and designers in estimating C.O.'s, 4) estimators in the checking of detailed estimates, 5) production managers in cost control of the work in progress, and 6) for contractors in last minute bid preparation(Siqueira 1999).

Parametric cost estimate can be conducted using three main techniques; traditional technique, computer based technique and artificial intelligence based technique. In the following section all these techniques will be discussed.

2.4.4.1 Traditional Techniques

General forecasting techniques are basically quantitative approaches that have been in use for the past half century. They can be either deterministic or stochastic. Examples of deterministic methods are regression methods (linear regression and multiple regressions), econometric models, moving average methods and exponential smoothing methods. Examples of stochastic methods are the maximum likelihood method, Box-Jenkins models, and probability weighted moment (L-moment). In construction, deterministic cost estimation models (referred to as parametric models) have traditionally been in wide use by many researchers due to their simple formulations. In these techniques, historical data are used to develop cost relationships based solely on statistical analysis. They have been used to estimate one characteristic of a system, usually its cost, from other performance characteristics of the system (Ayed 1997).

In the traditional technique of parametric analysis, the main parametric equation has to be defined with its associated independent variables before any analysis can be performed. The variables in the parametric equation could be identified by highlighting those characteristics of the system under study that are most directly related to its cost. Then, a mathematical relationship for correlating data could be used to express this relationship. The most popular mathematical models are linearized forms of the arithmetic, logarithmic and semi-logarithmic equations. In the parametric cost estimating technique, there is one dependent variable which is the cost and two or more independent variables like size, location, capacity, time ....etc. The equations generally take one of the following three forms (Gwang 2004):

1) Linear Relationships: $Cost=a+bX_1+cX_2 + ...$
2) Logarithmic Relationships: $Log\ (Cost) =a+b\ Log\ X_1 + c\ log\ X_2+.........$
3) Exponential Relationships: $Cost=a+ b\ X_1^c + d\ X_2^e+.........$

Where a, b, c, d and e are constant and X_1, X_2, X_3, X_4, ..., X_n, are the performance characteristics of a system.

The best criterion for choosing a form of the cost estimating relationship is a good understanding of how costs vary with changes in the independent parameters. This is a difficult task that is based on the experience of the estimators involved and, as such, has been performed mainly on a trial and error basis.

A major disadvantage of the traditional techniques for parametric estimating is that the mathematical form has to be defined before any analysis can be performed to determine the actual cost function that best fits the historical data. Also, modeling the cost of a system as a function of a member of independent variables is not an easy task. This is due to the large number of variables present in the system under evaluation and the numerous interactions among them. Another drawback is the use of a single cost estimation relationship to all cost variables involved which often have different mathematical correlation with the cost of that system. These problems may explain the low accuracy and limited use of such parametric estimating technique based on mathematical analysis in the construction industry, and the need for a more accurate technique to solve the cost estimation problems (Gwang 2004).

Several research efforts have developed for parametric cost estimation models based on traditional estimating techniques. Some of these efforts assist in generating cost indexes or production rates in addition to traditional parametric analysis. Those efforts where mentioned in Ayed (1997) as following: “Al-Bani (1994) developed a concrete cost estimate model for small residential buildings. The research studied the inter-relationships between the different physical elements of a concrete structure such as footings and columns using mathematical expressions and formulas. In a different approach, Lopez (1993) conducted a study on forecasting construction costs in hyper-inflated economies. In this study, Mexican economic indicators were compared with those from the United States. As a result a new Mexican cost index was developed using Box and Jenkins models. Multiple regression analysis has also been used by Pantzeter (1993) to develop a methodology for modeling the cost and duration of concrete highway bridges. In this research, different projects were divided into five work categories and the cost of each category was modeled by applying statistical techniques. A similar effort was also conducted by Akeel (1989) in developing a database tool for statistically based construction cost estimating. This research utilized the multiple regression analysis to develop the cost estimating relationship".

2.4.4.2 Computer-Based-Techniques

During the recent years, the value of computer aided cost estimation has grown as a manufacturing tool. Computers can be employed to produce the cost estimates using the same techniques that have been developed for manual cost estimation. By using computers, better quality estimates or the same quality can be produced quicker and cheaper (Tunc 2003).

The application of computers in the cost estimation process offers many advantages to cost estimators. Basic advantages are the speed and the accuracy. Cost estimation often involves complex mathematical calculations and requires advanced mathematical techniques. The computer is fast and may relieve the estimator of most of the job of making numerous calculations, which involve chances for making errors. However, although the computer may be able to complete a job in a matter of seconds, the time required for developing and entering the inputs may take a time of several minutes or hours (Tunc 2003).

With the advantage of computers in the 1980's, commercial software systems for cost estimation have proliferated. Some of these systems, such as "Timberline", are fully integrated systems that allow estimating module to exchange data with other modules, such as CAD (Ayed 1997).

Also, Ayed (1997) stated that "The Commercial software packages such as 'Success System' and 'Design 4/Cost' are used at the early stage of a project when data is not sufficient to perform a detailed cost estimate". On the one hand, (Success) helps in providing a parametric as well as a detailed estimate of a project based on independent variables associated with each item of work. Also the software still needs user identification of how to correlate parameters affecting an item to its cost which is a difficult task. On the other hand, (Design 4/Cost) uses a historical database of previous projects executed in The United States. The software uses algorithms and mathematical relationships for regional adjustment and cost escalation inside the United States for up to year 2000. All estimates, however, are generated using the square feet building area of the project as the main parameter. The accuracy of the estimates, however, degrades if the new building area is not within a ±20% range, relative to the historical database (Ayed 1997).

It is important to emphasis that the computer is a tool to assist in performing of estimates and not a substitute for the estimator. Basically, the computer is a tool to complement the estimator. The computer program can be too automated for the benefit of the user, but it can never be a substitute for the ability of the estimator to develop reasonable and realistic inputs based on the judgment and experience. Although, computers cannot be used for the entire estimation process, they can relieve estimators of much of the hard work associated with routine, repetitive, and time consuming calculations (Tunc 2003).

Ayed (1997) briefed several surveys which have been conducted to identify the commercial cost estimation software, those surveys were as flows: ”Yau (1992) implemented an object-oriented methodology to develop a model for integrated scheduling and a cost estimate at the preliminary design stage of mid-rise building projects. Another study by Lee (1992) developed a cost estimating model that can provide detailed cost information on design alternatives at all stages of design for reinforced concrete buildings. Hollman (1994) introduced a parametric cost estimating system for buildings developed by the capital estimating department at Kodak Park. The system can give a rough cost estimate with ±5% accuracy for the total cost of buildings. The advantage of this system is the elimination of the need for detail or assembly level cost data. Also, the U.S. army Corps of Engineering (COE) has developed a software program that can be used for preparing a quantitative estimate from parametric models. The program is known as Control Estimate Generator. This program was used by Melin (1994) to demonstrate various aspects of parametric estimating and standard estimating software. The parametric models are designed to generate a detailed estimate from historical or past projects. The program uses a parametric approach to adjust the original quantities by parameters to generate new quantities for use in the proposed estimate".

Also, Tunc (2003) concluded some other efforts as the following: "Guler and Gokler (1990) studied on evaluating the early cost for plastic injection molds. In this study, final cost was separated into additive cost items, which are assigned to individual mold functional components. After inputting molding properties like geometrical dimensions, surface quality etc. the program outputs the total cost for the molds. Shaw (1997) studied on machining and metal cutting techniques and offer methods for time estimation in machining. Bouaziz (2001) studied on evaluating machining costs in plastic blowing molds. This study, which is based on classical machining formulas, is also convenient to apply on forging die machining cost estimation".

As opposed to the deterministic approaches described earlier, probabilistic techniques (e.g., Monte Carlo simulation) have recently been used for estimating. The benefit of such techniques is their ability to account for risks and model the cost components with high variability. The simulation considers statistical distribution for each cost variable under estimation and produces a value for the total cost estimate and its probability of occurrence. Despite the probabilistic advantage of the Monte Carlo simulation, it has two major limitations. First, statistical distributions for various cost components need to be established in advance. Second, if the cost variables are not independent, their correlation should be accounted (Tunc 2003).

It is clear from the previous discussion that computers are very fast and accurate when carrying out the extensive computations needed for cost estimating. They are, however, traditionally poor at the intuitive aspects that require the integration of experience and making decisions that cannot be clearly defined in mathematical form. Therefore, they fail to adequately model the essential part of estimating that is based on experience and analogy with previous situations. This has motivated the application of non-traditional estimating techniques based on artificial intelligence (AI) in this domain.

2.4.4.3 Artificial Intelligence-Based-Techniques

Artificial Intelligence (AI) has been a rapidly growing field of computer science that has direct applications in the construction industry. The term (AI) is applied to those fields of computer science attempt to simulate human intelligence. Expert Systems and Neural Networks are among current Artificial intelligence research areas represent a great promises. On one hand, Expert Systems attempt to model the intelligent reasoning and problem-solving capabilities of the human brain. They use rules in the form of (IF, ..., THEN) to explain how they arrived to reliable decisions. Several expert-system applications were used in the field of construction engineering and management; two of these are specialized in cost estimating at early stage of the project.

On the other hand, the Neural Networks technique presents itself as a new approach of computation and decision making that may potentially resolve some of the major drawbacks of traditional estimating techniques. It holds a great promise for rendering the parametric method of cost estimating in a reliable and reasonably accurate way to prepare cost estimates.

In addition, Neural Networks take a different approach than traditional (AI) techniques such as Expert Systems. Neural Networks attempt to replicate the mechanism by which the human brain manipulates data and reaches decisions (Tunc 2003).

Several researchers have used neural networks as a tool for estimating costs for the earlier stage of project development. Hegazy and Ayed (1998) developed a cost-estimating model for highway projects using neural network approach and effectively manage construction cost data. A simple neural network simulation has been developed in a spreadsheet format. Simplex optimization and genetic algorithms techniques were used to determine network weights. Setyawati et al (2002) developed a neural network for cost estimation and suggested regression analysis with combined methods based on percentage errors for obtaining the appropriate linear regression which describe the artificial neural network models for cost estimating. Gwang et al

(2004) examined different methods of cost estimation models in the early stage of building construction projects such as multiple regression analysis and neural networks. They concluded that neural networks performed best prediction accuracy. Murat et al (2004) developed a cost estimation model for building based on the structural system for the early design process. They suggested that their model establishes a methodology that can provide an economical and rapid means of cost estimating for the structural system of future building design process. They argued that neural networks are capable to reduce the uncertainties of estimate for a structural system of building and the accuracy of the model developed was 93% level.

Jamshid (2005) also examined cost estimation for highway projects by artificial neural network and argues that neural network approach might cope even with noisy data or imprecise data. He reported that artificial neural network could be an appropriate tool to help solve problems which comes from a number of uncertainties such as cost estimation at the conceptual phase. Because of back-propagation, neural network has good nonlinear approach ability and higher prediction accuracy, thus used for cost prediction.

Bouabaz et al. (2008) developed an Artificial Neural Network to estimate the total cost of bridges repair and maintenance with high accuracy in developing countries. Cost and design data for two categories of repair bridges were used for training and testing our neural network model, with only three main parameters used in estimating the total cost of repairing bridges. An accuracy of 96% was achieved.

Another attempt done by Lekan (2011) to develop Artificial Neural Network for cost estimating of total cost for building projects. The efficiency of the developed model developed is found to be reasonably high, with relative average efficiency of 0.763 and coefficient of performance of 1.311, which is adjudged good. Also, Arafa et al. (2011) developed an Artificial Neural Networks model to estimate the cost of building construction projects at the early stage. The study was used a database of 71 building projects collected from the construction industry of the Gaza strip, Palestine. The results obtained from the model indicated that neural networks reasonably succeeded in predicting the early stage cost estimation of buildings using basic information of the projects and without the need for a more detailed design. Hosny (2011) developed an Artificial Neural Networks model that can materially help construction planners in the accurate determination of the expected time contingency of any future building projects. One hundred trials were applied to identify the best structure of the proposed model. The developed model predicts the time contingence with accuracy 0.0% to 7.5%. Also Asal (2014) developed an Artificial Neural Networks model that can be used to estimate the cost variance for any type of construction building projects in Egypt. The study was used a data base enclosing various types of building projects. The accuracy of the developed model was -23% to +23%.

2.5 Procedures of Parametric Estimating

The development of a new parametric cost estimating model is complicated process, starts with thoughts and continues till turning out the desired predict cost with acceptable accuracy. Accordingly, this process requires to be simplified, the best way to facilitate any process is to create a course of actions that enclosing the entire fractions of this process. Therefore, the development of any parametric cost estimating model commonly follows a certain steps in order to facilitate this process. Thus, six steps were conceived as the backbone for the procedures needed to produce the proposed output. These steps were illustrated in figure (2.2).

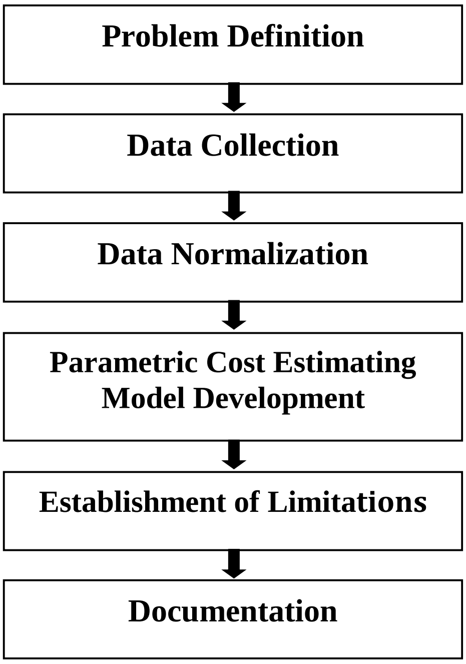


Figure (2.2): Parametric Cost Estimating Procedure (Murat 2004)

The first step is the definition of the problem. The problem definition is the first step in any scientific method. It encompasses the determination of the objectives and scope of this effort (Murat 2004).

The second step is the data collection. Parametric estimating requires an extensive database, where historical records are extremely important. The design or engineering parameters that drive parametric cost estimates are developed from historical cost databases. Data collection can be regarded as the most crucial step. Without sufficient relevant data, the parametric estimating cannot be successfully implemented.

The third step is the normalization of the data. This process ensures that every individual record in the database is on the same base. Typically, in construction project, the cost data for each project must be adjusted for time and location differences. This step is important and must be performed before further analysis can be made to the data (Murat 2004).

The fourth step is the parametric cost estimating model development. It involves the determination of interdependencies of the variables to be used in the model and the derivation of the cost estimating relationships. Cost estimating relationships are mathematical models or graphs that estimate cost. CERs are statistical models that characterize the cost of a project as a function of one or more independent variables or parameters. Rules of thumb and the unit-quantity method are not recognized as CERs (Phaobunjong 2002). The basis for selecting the parameters for use in the model should be more than just the statistical validity, but the inclusion of the parameters should also be based on the logical and theoretical considerations.

Traditionally, cost estimating relationships are developed by applying regression analysis to historical project information (Hegazy et al. 1998). Regression analysis is a good method of determining the relationship between the parameters and cost, and determining the exact mathematical form of the model.

The fifth step of the parametric estimating procedure is the establishment of model limitations. The model is generally developed from a limited data set, therefore valid only for the ranges of variables used in the model. Interpolation is, of course, more accurately done, and extrapolation must always be done carefully and the range over which the estimate is valid should be known. Extrapolation beyond this range will produce unverified results.

The sixth and final step is the documentation of the model development process. The assumptions and limitations of the model must be properly documented to facilitate the successful model implementations. Notes should also be recorded for any uncertainties in the data and in the estimate. Information and meanings of the terms used in the data collection and the model development must be documented along with all the calculation methods (Murat 2004).

2.6 Cost Drivers for Parametric Estimating of Construction Projects

The most critical approach for developing parametric cost estimating model is the determination of cost related drivers or cost independent variables, especially for a certain type of buildings that are not widely used such as sterile buildings. The accuracy of the model will be directly affected by these drivers; however, the available data for estimating the cost, at the early stage, are limited or not even exist.

Many papers have determined the cost related drivers in try to develop adequate models for parametric cost estimation of construction projects, these efforts was used a different data pools for a various types of buildings. The first study done by Emsley et al. (2002) was based on a data pool of 288 properties. They worked with regression analyses as well as with neural networks to compare the two statistical methods. Up to 41 independent variables were included in the developed models, and in the best case the developed model supplied a Mean Absolute Percentage of Error (MAPE) of 17%. The practical application of the models was made more difficult due to this comparatively large error. In addition, the necessary input (41 independent Variables) was extensive, making it difficult to apply in early design stages.

Another effort done by Wheaton and Simonton (2007) developed hedonic cost models for residential and office properties. Their study was based on data available for over 60,000 properties (with 42,000 residential properties) and primarily concerned six American markets. The focus of their work was on the “true” trends of the cost during a period of 35 years. In addition, they analyzed the correlation between costs and building activity. The supply of cost indicators and their drivers was not central to their study. However, the present study can work with the developed semi-log regression models and their five independent variables.

Stoy et al. (2008) have used pre-mentioned studies in addition to other old studies to develop a list of cost drivers. Table (2.1) presents the studies and the relevant cost drivers that were collected by Stoy et al. (2008).

Table (2.1) Related Studies and Their Relevant Cost Drivers [Adapted from Stoy et al. (2008)]

Study	Data Pool	Properties of the Model	Cost Drivers (Extract)
Bowlby et al. (1986)	157,855 projects	Regression	• Number of stories • Total area of building • Metropolitan or rural location • End use of building • Building framing types • Region of U.S. where building is located
Thalmann (1998)	15 properties (residential)	Regression (semilog)	• Usable floor area • Proportion of external wall areas underground • Proportion of openings in external wall areas • Year of construction
Elhag and Boussabaine (1998)	30 properties (schools)	Neural networks	• Type of building (primary & secondary school) • Gross floor area • Number of stories • Project duration
Emsley et al. (2002)	288 properties	Neural networks and regression	• Project strategic variables (e.g., type of contract) • Site-related variables (e.g., topography) • Design-related variables (e.g., gross internal floor area)
Elhag et al. (2005)	Literature and interviews	----	• Client characteristics • Consultant and design parameters • Contractor attributes • Project characteristics • Contract procedures and procurement methods • External factors and market conditions
Li et al. (2005)	30 properties (office)	Regression (linear)	• Frame type • Floor area • Building height • Average floor area
Love et al. (2005)	161 properties	Regression (log-log)	• Duration (time-cost relationship) • Gross floor area
Wheaton and Simonton (2005)	42,340 residential properties 18,469 office properties	Regression (semilog)	• Number of stories • Absolute size • Number of units • Frame type • Year of construction
Asal (2014)	42,340 residential properties 18,469 office properties	Regression (semilog)	• Total built-up area • Consultant fees • Contractor overhead • Type and skills of contractor • Cement and reinforcement prices • Site topography

As shown in the above presentation, numerous building parameters are suggested by the pre-mentioned studies. However, it is evident that the parametric cost estimating methods tends to use general data at a macro level, where detailed data are not available.

This thesis will use the previous efforts, of determining the cost drivers, in addition to the questionnaire technique to designate the required cost drivers that needed to provide a parametric cost estimating model for such sterile type of buildings. This study will focus only on determining the macro level data used to develop a reliable parametric cost estimating model, this approach will be extensively discussed and presented in further chapter (4).

2.7 Conclusion

Heretofore, the importance of the conceptual Cost estimating was discussed in this chapter. In addition, the old and new techniques used to perform parametric cost estimating were demonstrated. Accordingly, it was discovered that using of Artificial Intelligence (AI), especially Artificial Neural Network (ANN) approach, may potentially resolve some of the major drawbacks of traditional estimating techniques. It holds a great promise to enhance the reliability and reasonability of parametric cost estimating method.

Moreover, factors affecting the total construction cost of various types of building projects were identified from previous researches, these factors are as the following: (1) project location, (2) desired completion time for the project, (3)site topography, (4) accumulative built-up area, (5) other supplementary buildings, (6) desired structural system, (7) consultant fees, (8) desired level of contractor's prequalification, (9) contractor overhead, (10) need for special contractor(s), (11) reinforcement price, (12) cement price, (13) labor price, (14) inflation.

CHAPTER 3

ARTIFICIAL NEURAL NETWORK

CHAPTER 3

ARTIFICIAL NEURAL NETWORK

3.1 Introduction

Artificial Neural Networks (ANNs) is the most recent approach among the artificial inelegance field. It presents itself as a new approach of computation and decision making. Artificial Neural are not new; they were introduced half a century ago. ANNs**,** however, have received considerable attention in recent years, with the rapidly increasing computational power and concurrently declining cost of computers (Lekan 2011).

Firstly, this chapter will present an extensive literature about the artificial neural networks to illustrate its advantages and drawbacks, thereafter, the study will focus on using the genetic algorithms as a training algorithm for the neural networks model. Finally, this chapter will represent some of previous efforts that conducted artificial neural networks in field of construction projects.

3.2 Concept

Artificial neural networks try to reproduce the generalization abilities of a human neural system. Neural Networks are particularly effective for complex estimating problems where the relationship between the variables cannot be expressed by a simple mathematical relationship. They are computer programs simulating the biological structure of the human brain which consists of tens of thousands of highly interconnected computing units called neurons. An Artificial Neural Network can be constructed to simulate the action of a human expert in a complicated decision situation.

The most commonly used type of ANNs, known as the feed-forward or back propagation. This type and similar architecture such as general regression neural networks (GRNN) are most suited for pattern recognition and forecasting class of problems (Hosny 2011)

3.3 Structure

The architecture of ANNs contains several elements such as: input, connecting weights, summation, activation function and output. Input and output elements are called neurons. An artificial neuron is a unit processing element contains a single perceptron to compute the output of network by forming linear combination of activation functions and weights (Attal 2010).

A Neural Network is constructed by arranging several processing units in a number of layers figure (3.1) shows the general structure of Artificial Neural Network. The output of a single layer provides the input to the subsequent layer and the strength of the output is determined by the connection weights between the processing units of two adjacent layers.

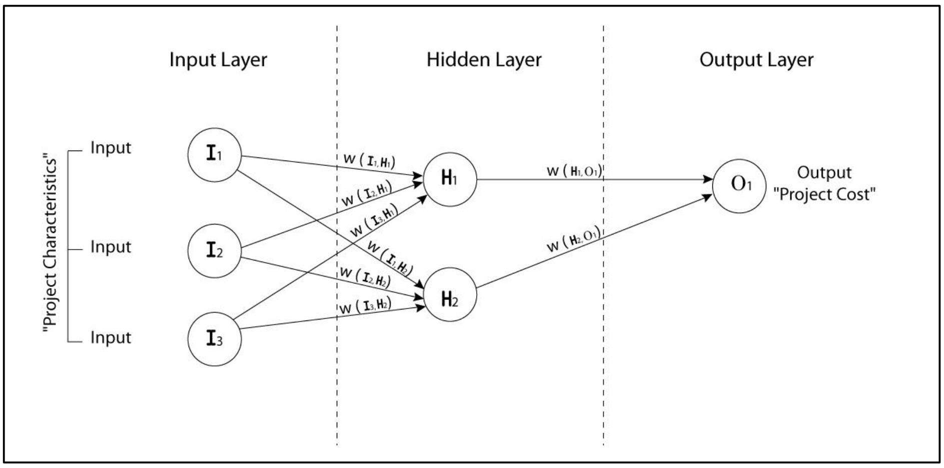


Figure (3.1): Artificial Neural Network Structure

Where I_1, I_2, ., I_n are an input data of the project characteristics, and H_1, H_2, .. H_n are the hidden nodes, and O_1 is the output of the network what represent the solution of the problem under study, and $W_{(Ix.Hx)}$ and $W_{(Hx.Ox)}$ are the connecting weights

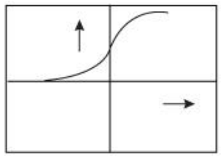
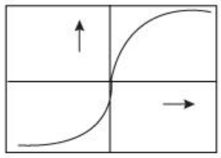
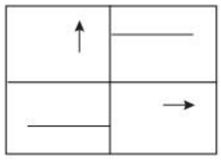
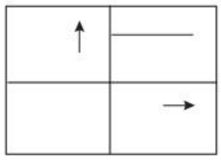
In terms of the structure of ANNs, one or more activation functions are used to activate or interpret data inside each neuron from the inlet node to outlet one. There are several types of activation functions used in ANN, the most common used activation functions in the field of engineering, specially construction engineering, are mathematical functions (Attal 2010), the most common mathematical activation functions used in ANNs to interpret the data were presented in table (3.1).

Activation functions for hidden units are needed to introduce nonlinearity into the network. Without nonlinearity, hidden units would not make networks more powerful than just plain Neurons (which do not have any hidden units, just input and output units). However, it is the nonlinearity (i.e, the capability to represent nonlinear functions) that makes multilayer networks so powerful. Almost any nonlinear function does the job, for back-propagation learning, the activation function must be differentiable, and it helps if the function is bounded; the sigmoidal functions such as logistic (sigmoid) and tanh function are the most common choices. Functions such as tanh that produces both positive and negative values tend to yield faster training than functions that produce only positive values such as logistic (sigmoid), because of better numerical conditioning (Salam et al. 2012).

For hidden units, sigmoid activation functions are usually preferable more than threshold activation functions. Networks with threshold units are difficult to train because the error function is stepwise constant, hence the gradient either does not exist or is zero, making it impossible to use back-propagation or more efficient gradient-based training methods. Even for training methods that do not use gradients methods, such as simulated annealing and genetic algorithms, sigmoid units are easier to train than threshold units. With sigmoid units, a small change in the weights will usually produce a change in the outputs, which makes it possible to tell whether that change in the weights is good or bad. With threshold units, a small change in the weights will often produce no change in the outputs (Attal 2010).

For the output units, an activation function should be chosen to suit the distribution of the target values (Salam et al. 2012).

Table (3.1) Mathematical Activation Functions Used in ANNs Artificial [Adapted from Attal (2010)]

(1) Sigmoid function	$\frac{1}{(1 + e^{-x})}$	
(2) Tanh function	$\frac{(e^{x} - e^{-x})}{(e^{x} + e^{-x})}$	
(3) Signum function	+1 , if $x > 0$ -1 , if $x \le 0$	
(4) Step function	+1 , if $x > 0$ 0 , if $x \le 0$	

From previous discussion we can discover that one of the most important characteristics of Neural Networks is their ability to learn from a set of examples to detect by themselves the relationships that link inputs to outputs.

For training the model, all historical data are categorized into two sets of data, one called training data and the other called testing data. The commonly used distribution ratio of historical data are 20% of cases for the training and 80% are for testing the results (Bouabaz, et al. 2008).

The training input data is presented to the input layer, normalized, and multiplied by its connection weight. Calculated weighed inputs are summed at each node to produce an activation value, which is then modified by a selected transfer function. The output of the neurons in the hidden layer(s) will then be multiplied by the respective weights, this product summed up, and a new activation value generated. This new value will then be modified by a transfer function, and an output for that processing element is calculated

During training, both the inputs (representing problem parameters) and outputs (representing the solutions) are presented to the network normally for thousands of cycles. At the end of each cycle, or iteration, the network evaluates the error between the desired output and actual output. Then uses this error to modify the connection weights according to the training algorithms used.

One of the most common training algorithms is Back propagation. For a certain number of learning cycles the weights are continuously adjusted until the deviations from the desired outputs are minimized. After training, the network can be used to predict the solution for a new case not used in training (Salam et al. 2012).

3.4 Genetic Algorithms Based Software

Genetic algorithms based software seek to solve optimization problems using the method of evolution, specifically survival of the fittest. The theory of GAs is that a population of a certain species will, after many generations, adapt to live better in its environment. For example, if the species is an animal that lives mainly in a swampy area, it may eventually evolve with webbed feet. GAs solves optimization problem in the same fashion. It will create a population of possible solutions to the problem. It solves the problem by allowing the less fit individuals in the population to die and selectively breeding the fittest individuals (the ones that solve the problem best) of the population. This process is called selection, as in selection of the fittest. GAs takes two fit individuals and mates them (a process called crossover). The offspring of the mated pair will receive some of the characteristics of the mother, and some of the father. Offspring often have some slight abnormalities called mutations. After GAs mates fit individuals and mutates some, the population undergoes a generation change. The population will then consist of off springs plus a few of the old individuals which was allowed to survive to the next generation because they are the most fit in the population and it is wanted to keep them breeding. The fittest individuals are called elite individuals. After dozen or even hundreds of generations, a population eventually emerges wherein the individuals will solve the problem very well. In fact, the fittest elite individual will be an optimum or close to optimum solution (Lekan 2011).

3.5 Problems with Neural Networks

Despite the good performance of Neural Networks in the previous studies, the process of developing and implementing Neural Networks to parametric cost estimation has a number of problems associated with it. First, designing the network architecture and setting its parameters is not a straight-forward approach; it actually requires some trial and error process. A large amount of time must be spent determining the best network architecture and the network parameters that best fit the application under consideration (Lekan 2011).

Second, the learning algorithms, such as Back-propagation, require optimization of the network training in order to achieve adequate generalization, otherwise the memorization capability will be used by the network to memorize the ideal results. Such problem can be avoided easily at training stage by prohibition of network error to be equal to zero, in fact zero network error means memorized network.

Also, there is no explicit set of rules to determine whether a given learning algorithm is suitable for a particular application or not. In addition, the little user control over training and the final status of network weights have contributed to their black box perception.

Also, it should be noted that Neural Networks do not perform well with applications where precise numerical computations are required, like detailed estimating and cost control (Lekan 2011).

3.6 Applications in Cost Estimation

Several researchers have used neural networks as a tool for estimating costs for the earlier stage of project development. Hegazy and Ayed (1998), used a neural network approach to manage construction cost data and developed a parametric cost estimating for highway projects. They introduced two alternative techniques to train network's weights: simplex optimization (Excel's inherent solver function), and genetic algorithms, which is a flexible and adaptable model for estimating highways projects by using a spreadsheet simulation. Adeli et al. (1998), formulated a regularization neural network to estimate highway construction costs and indicated that the model were very noisy and this noise results from many unpredictable factors related to human judgment, such as random market fluctuations, and weather conditions. They concluded that the model is successful in introducing a number of attributes to make it more reliable and predictable.

Also, Graza et al. (1995) compared the results of Neural Networks with those of regression models for predicting the material cost of carbon steel pipes. Ten sets of cost data were used to train the Neural Networks and six sets were used in testing. The costs for these 6 sets were estimated using three methods (linear regression, nonlinear regression and Neural Networks). The mean square error was calculated for each model. It was found that Neural Networks produced the lowest mean square error compared with the other two models. The accuracy level was between 66.8% and 77.96%. This study proves that the Neural Networks approach could be used to resolve some of the major drawbacks of the regression-based parametric estimation.

Gwang et al. (2004) examined different methods of cost estimation models in the early stage of building construction projects such as multiple regression analysis and neural networks. They concluded that neural networks performed best prediction accuracy. Murat et al (2004) developed a cost estimation model for building based on the structural system for the early design process. They suggested that their model establishes a methodology that can provide an economical and rapid means of cost estimating for the structural system of future building design process. They argued that neural networks are capable to reduce the uncertainties of estimate for a structural system of building and the accuracy of the model developed was 93% level.

Another study was conducted by Mckim (2005) worked on application neural networks in identification/estimating of risk. The study used neural network to predict percentage change of the final cost from the rough estimated costs as an index of risk measurement. 150-sampled construction projects were used for the study, with parameters such as final costs, project size and estimated costs used as neural network system input data. The study obtained percentage change in cost of the actual estimated cost to final cost and submitted that it can be used as a risk indicator for construction projects. The study further stated that, the mean variance obtained from neural network results was less than variance of mean overrun cost; and that neural network technique produces good result only in case of rough estimation. The study recommends this approach when risk is to be determined at early stage of construction project. The reviewed study scope did not include cost prediction of building works, it centered on risk measurement as well not on producing a model for cost prediction. To this end therefore, in this study, a model that can be used in building cost predicting was generated which can be used at early, middle and latter stages of building works.

Creese et al. (2005) studied the cost estimating of timber bridges using neural networks. The study deployed neural network and the output was simulated with output from regression analysis in order to determine which approach has least mean square error (r-square values). The study used cost parameters of the timber bridge such as volume of the web, volume of the decks, wooden flange and bridge weight as neural network input data. The study with neural network indicated that, the r square values using neural network system were greater than when regression analysis was used. So also, according to the researchers‘ submission, in estimating cost of timber bridges, the models with 3 – input variables gave the least error; and Neural network systems give little variance to the actual cost from expected cost. Neural network however decreases in margin error as input variable increases, the large the sample, better the accuracy; to this end therefore sample of 500 magnitudes was used in the research under review. Expectation of this research work in this regard is that, there is assurance of the fact that the variance of actual construction cost from the initially budgeted cost will be negligible since larger sample was used.

Also, Jamshid (2005) examined cost estimation for highway projects by artificial neural network and argues that neural network approach might cope even with noisy data or imprecise data. They reported that artificial neural network could be an appropriate tool to help solve problems which comes from a number of uncertainties such as cost estimation at the conceptual phase. Because back-propagation neural network has good nonlinear approach ability and higher prediction accuracy it has been used for cost prediction.

Emsley et al. (2007) suggested that procurement routes cannot be isolated from cost significant variables in a building project. Therefore, Al Tabtabai et al., (2008) developed a neural network model that could be used to estimate the percentage increase in the cost of a typical highway project from a baseline reference estimate such as environmental and project specific factors. Their model generates a mean absolute percentage error of 8.1%.

Meijer (2009) utilized neural networks for circuit modeling. Neural network was adapted in device and circuit modeling. The study generated a semi –automatic modeling path that could be used for device and circuit modeling. Also, Meijer (2009) carried out a research work on the adaptation of generalized feed forward Networks; the study developed analogue modeling of continuous and dynamic non-linear multidimensional system for simulation. So also, the effect of using second order differential equations on feed forward networks was determined. The study identified among other things, the distribution of time steps that allows for smaller time steps during steep signal transaction in transaction feed forward networks. The study stated further, that, the use of second order differential equation for each neuron allows for efficient mapping to a general neural network formation. To this end, this research work developed a neural network based building project cost predicting model with genetic algorithm, using cost centers from building project document of completed works. The cost centers were adjusted with current price index. So also, a total sample of 500 magnitudes was used in this study for model development, validity and stability.

Another effort done by Hosny (2011) developed an Artificial Neural Networks model that can materially help construction planners in the accurate determination of the expected time contingency of any future building projects. One hundred trials were applied to identify the best structure of the proposed model. The developed model predicts the time contingence with accuracy 0.0% to 7.5%. Also, Asal (2014) developed an Artificial Neural Networks model that can be used to estimate the cost variance for any type of construction building projects in Egypt. The study was used a data base enclosing various types of building projects. The accuracy of the developed model was -23% to +23%.

3.7 Conclusion

Artificial Neural Networks have been developed at the recent decades. Artificial Neural Networks are particularly effective for solving complex problems, such as cost estimating problems, where the relationship between the variables cannot be expressed by a simple mathematical relationship.

The most important characteristics of Neural Networks is their ability to learn from a set of examples to detect by themselves the relationships that link inputs to outputs, this attribute expresses the capability of Artificial Neural Network not only to manipulate the historical data as human brain, but also to solve complicated problems by searching for the optimal or near optimal solution by using one of the evolutionary learning algorithms. In addition, several conclusions are derived based on this literature survey:

1. Back propagation is the most widely used Neural Network training algorithms especially for parametric cost estimation.
2. Special attention has to be focused on identifying the problem attributes and accordingly the relevant independent factors.

Special attention has to be focused on improving Neural Networks generalization either by using appropriate testing and verification or the use of optimization technique such as GAs.

CHAPTER 4

DATA COLLECTION AND ANALYSIS

CHAPTER 4

DATA COLLECTION AND ANALYSIS

4.1 Introduction

This Chapter tends to identify the main factors which are expected to have an effect on projects' cost in the Egyptian construction industrial field. Identifying these factors can help to enhance the accuracy of the cost estimating process, especially at the early stages of project life cycle. These factors will be identified based on the previous literature review in chapter (2) - table (2.1), and limited interviews with experts in field of pharmaceutical and food industrial projects.

This Chapter contains three sections. The first section will be oriented to the process of building the questionnaire and collecting results from all stakeholders. The second section provides some analytical calculation and logical discussion regarding the results of the questionnaire and the effect of these factors on project cost estimating and thereof the determination of the cost indicators to be considered in the modeling process. Finally, the third section will be specified for gathering the required data from previous projects.

4.2 Objective

The main objective of this chapter is to identify cost indicators for the cost estimating process for industrial construction projects specifically industrial projects encompassing sterile buildings, through discussing how the data collection was carried out, such as the questionnaire structure, the recipient stakeholders, the sample size and how the survey data was analyzed, leading to identify the important factors affecting the cost estimating of such types of projects. These factors will be ranked from high important to low important, identifying the important cost factors then studied to determine the most significant cost drivers (Cost Indicators) that will be used to develop the Model.

4.3 Methodology

This thesis uses a set of steps in order to facilitate the process of identifying the most significant cost drivers (Cost Indicators). Firstly, the questionnaire will be built using two components, first one is the factors concluded at the literature review from the previous chapter, and the second one is limited interviews with some experts in order to add additional factors that may be important in cost estimating process for such types of projects.

Secondly, the questionnaire will be sent to experts such as the owners, project managers, consultants, and contractors, in an attempt to cover wide points of view from all stakeholders for such types of projects. Then, the response from recipients will be gathered and summarized.

Thirdly, the gathered data will be analyzed then ranked from high important to low important through their importance index, in order to identify the important cost drivers, then grouped by logical relations to conclude the cost indicators to be used in building the proposed cost estimating model.

4.4 Questionnaire

4.4.1 Definition

A questionnaire is a research instrument consisting of a series of questions and other prompts for the purpose of gathering information from respondents. Although they are often designed for statistical analysis of the responses, this is not always the case.

Questionnaires have advantages over some other types of surveys in that they are cheap, do not require as much effort from the questioner surveys as verbal or telephone surveys, and often have standardized answers that make it simple to compile data. However, such standardized answers may frustrate users.

Questionnaires are also sharply limited by the fact that respondents must be able to read the questions and respond to them. Thus, for some demographic groups conducting a survey by questionnaire may not be practical.

4.4.2 Questionnaire Design

In this section a questionnaire survey was designed using a set of steps as follows:

I. Identifying Factors

The first set of factors was previously concluded in the literature review done in chapter (2). Fourteen cost factors were predetermined as shown in table (4.1).

Table (4.1): Cost Factors Collected From Literature Review

Factors Collected From the Literature Review
1. Project Location
2. Desired Completion Time for the Project
3. Site Topography
4. Accumulative Built-up Area
5. Other Supplementary Buildings(W. Tank, Administration, Warehouse,…etc)
6. Desired Structural System
7. Consultant Fees
8. Desired Level of Contractor's Prequalification
9. Contractor Overhead
10. Need for Special Contractor(s)
11. Reinforcement Price
12. Cement Price
13. Labor Price
14. Inflation

Moreover, limited interviews were conducted with four expert engineers in the field of pharmaceutical and food industrial projects, the interviews were restricted to projects managers and cost specialist with +25 years of experience. These interviews were done to determine additional factors regarding experts' judgments, especially for that such type of projects they expert in. As a result, twenty three new cost factors were derived from those interviews. Table (4.2) represents the additional factors derived from the interviews.

Table (4.2): Cost Factors Derived From Interviews With Experts

Factors Derived From Interviews
1. Site Accessibility
2. Site Constraints
3. Subsistence of Time Constrains
4. Owner Requirements for Bid Packaging for Multiple Contractors
5. Environmental Impact Assurance System Requirements
6. Applying Safety System During Construction
7. Buildings Closeness (Attached, Semi-attached or Separated)
8. Accumulative Sterile Areas (Total Area)
9. Structural Design Loads
10. Geotechnical Nature of Soil
11. Desired HVAC System
12. Desired Firefighting System
13. stainless steel price
14. % of Imported Material
15. Availability of Required Power
16. Target Market (Regional or International)
17. Type of Products and Type of Production Method
18. Special Finishing Required for Sterile Areas
19. Additional Requirements for Structural System Regarding Sterile Manufacturing
20. Additional Requirements for HVAC System Regarding Sterile Manufacturing
21. Industrial Safety Requirements (Firefighting, Fire Alarm, .. etc)
22. Currency Exchange Rate
23. International Insurances (if any)

II. Categorization of Factors

Next step is grouping the collected cost factors in categories according to the type and source of each factor. The main purpose of this step is enhancing recipient understanding to these factors. Categorization of factors was done through studying the source and type of each factor. Five categories were conceived as follows:

A. General Factors

General Factors represent the contributory factors related to the initial Project Data, Constraints and owner requirements

B. Engineering Related Factors

Engineering related factors represents the technical factors that can be calculated, assumed or demonstrated by numbers

C. Resources Related Factors

Resources factors are the factors that are directly related to the required amount of resources, their skill level and/or price

D. Special Factors

Special Factors are the mandatory factors related to the local and/or international regulations and requirements for factory production

E. External Factors

External factors represent the factors related to Local and international economy status

Thereafter, the collected cost factors were sorted into the pre-mentioned categories as shown in table (4.3)

Table (4.3): Categorization of Factors

Item	Category	Factors
A	**General Factors**	1. Project location 2. Site accessibility 3. Site Constraints 4. Desired completion time for the project 5. Subsistence of time constrains 6. Owner requirements for bid packaging for multiple Contractors 7. Environmental impact assurance system requirements 8. Applying safety system during construction
B	**Engineering Related Factors**	1. Site topography 2. Accumulative built-up area 3. Buildings closeness (attached, semi-attached or separated) 4. Other supplementary buildings(W. tank, administration, warehouse,… etc) 5. Accumulative Sterile Areas (total area) 6. Structural design loads 7. Geotechnical nature of soil 8. Desired structural system 9. Desired HVAC system 10. Desired Firefighting system
C	**Resources Related Factors**	1. Consultant fees 2. Desired level of contractor's prequalification 3. Contractor overhead 4. Need for special contractor(s) 5. Reinforcement price 6. Cement price 7. stainless Steel price 8. Labor price 9. % of imported material 10. Availability of required power

Item	Category	Factors
D	**Special Factors**	1. Target market (regional or international) 2. Type of products and type of production method 3. Special finishing required for sterile areas 4. Additional requirements for structural system regarding sterile manufacturing 5. Additional requirements for HVAC system regarding sterile manufacturing 6. Industrial safety requirements (firefighting, fire alarm, .. etc)
E	**External Factors**	1. Currency exchange rate 2. Inflation 3. International insurances (if any)

The percentage weight for the General Factors is 22%, the Engineering Related Factors is 27%, the Resources Related Factors is 27%, the Special Factors is 16% and the External Factors is 8%. Figure (4.1) illustrates the percentage of each category.

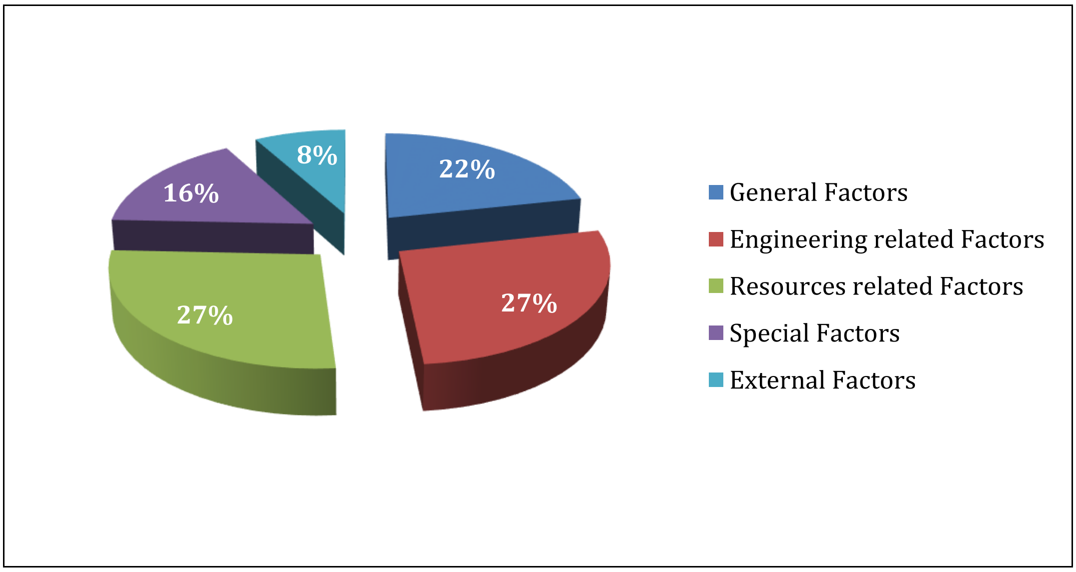


Figure (4.1): Categories Percentage

III. Questionnaire Scale

The questionnaire was based on a qualitative five-point scale to facilitate the answering process to recipients. Each recipient was asked to determine both "the impact" and "the degree of existence" for each factor, in accordance with their experience-based judgment. This scale varies from "very low to very high" as shown in table (4.4).

Where, "very low" means that this factor has a very weak effect on cost and the factor existed scarcely or changeability rarely happen, and "low" means that this factor has a weak effect on cost but might be existed or changed. However, "moderate" means that this factor has a medium effect on cost and has 50%/50% to be existed or changed, whilst "high" means that this factor has a huge effect on cost and frequently existed and changeable. Finally, "very high" means that this factor has a drastic effect on cost and there is no doubt of its existence nor changeability.

Table (4.4): Qualitative Scale Used in Questionnaire

Degree	Very Low	Low	Moderate	High	Very High
Abbreviation	V.L	L.	M.	H.	V.H

Thereafter, in the questionnaire analysis, the previous qualitative scale will be converted into a quantitative scale varying from 1 to 5 according to the qualitative scale respectively. Table (4.5) illustrates the qualitative scale and equivalent quantitative value.

Table (4.5): Quantitative Scale Used to Interpret Questionnaire

Degree	Very Low	Low	Moderate	High	Very High
Abbreviation	V.L	L.	M.	H.	V.H
Abbreviation	1	2	3	4	5

IV. Questionnaire Form

From the previous steps, a pilot questionnaire form was reached, this pilot form was sent to a limited number of experts to collect their comments and suggestions to reach a final form. The following form represents the final questionnaire. Table (4.6) represents the questionnaire form.

Table (4.6): The Questionnaire Form

Cost Estimation – Questionnaire
Definition Cost Factors are the contributory factors affecting the total cost of projects

No.	Cost Factors	Impact					Degree of Existence				
		V.L	L	M	H	V.H	V.L	L	M	H	V.H
A	**General Factors** General Factors represent the contributory factors related to the initial Project Data, Constraints and owner requirements										
1	Project location										
2	Site accessibility										
3	Site Constraints										
4	Desired completion time for the project										
5	Subsistence of time constrains										
6	Owner requirements for bid packaging for multiple contractors										
7	Environmental impact assurance system requirements										
8	Applying safety system during construction										
B	**Engineering Related Factors** Engineering related factors represents the technical factors that can be calculated, assumed or demonstrated by numbers										
1	Site topography										
2	Accumulative built-up area										
3	Buildings closeness (attached,										

No.	Cost Factors	Impact					Degree of Existence				
		V.L	L	M	H	V.H	V.L	L	M	H	V.H
	semi-attached or separated)										
4	Other supplementary buildings (W. tank, administration, warehouse,.. etc)										
5	Accumulative Sterile Areas (total area)										
6	Structural design loads										
7	Geotechnical nature of soil										
8	Desired structural system										
9	Desired HVAC system										
10	Desired Firefighting system										
C	**Resources related Factors** Resources factors are the factors that are directly related to the required amount of resources, their skill level and/or price										
1	Consultant fees										
2	Desired level of contractor's prequalification										
3	Contractor overhead										
4	Need for special contractor(s)										
5	Reinforcement price										
6	Cement price										
7	stainless Steel price										
8	Labor price										
9	% of imported material										
10	Availability of required power										
D	**Special factors** Special Factors are the mandatory factors related to the local and/or international regulations and requirements for factory production										
1	Target market (regional or international)										
2	Type of products and type of production method										
3	Special finishing required for sterile areas										

No.	Cost Factors	Impact					Degree of Existence				
		V.L	L	M	H	V.H	V.L	L	M	H	V.H
4	Additional requirements for structural system regarding sterile manufacturing										
5	Additional requirements for HVAC system regarding sterile manufacturing										
6	Industrial safety requirements (firefighting, fire alarm, .. etc)										
E	**External Factors** External factors represents the factors related to Local and international economy status										
1	Currency exchange rate										
2	Inflation										
3	International insurances (if any)										

Recipient Data

Company Name:	
Recipient Name:	
Recipient Title:	
Recipient Experience:	5 ☐ 10 ☐ 15 ☐ 20 ☐ 25 ☐
Signature:	
Date:	

Key

	V.L	L	M	H	V.H
	Very Low	Low	Moderate	High	Very High

4.4.3 Questionnaire Sample Size

I. Population

The research population constituents are mainly the expert stakeholders in the field of construction for the sterile buildings. The targeted population was framed to determine the opinion of 500 expert engineers in such type of projects in an attempt to cover reliable points of view (Bartlett 2001).

II. Sample Size Calculation

This study used probability sampling technique to calculate the required sample size (questionnaire respondent). In addition, the confidence level was selected to be 95%, and the confidence interval to be 15%, the Sample Size was computed as per equation (4.1) (Bartlett 2001).

$$SS = \frac{Z^2 \text{ x } (p) \text{ x } (1 - p)}{C^2} \quad \text{.......} \quad (4.1)$$

Where SS = Sample Size, Z = Z-values (Cumulative Normal Probability) represent the probability that a sample will fall within a certain distribution, the equivalent Z-value for a 95 percent confidence level equals 1.96, P = Percentage of population picking a choice, expressed as decimal, in this research questionnaire (P) is equal to (20%) according to the number of answers (five answers) for each question, but since 50% is the critical case percentage in the calculation of sample size, the value used for P in the equation is 0.5, C = Confidence interval, expressed as decimal (0.15 = +/- 15 percentage points).Therefore the Sample Size equals:

$$SS = \frac{(1.96)^2 \text{ x } (0.5) \text{ x } (1 - 0.5)}{(0.15)^2} = 43$$

For population less than 50,000, sample size shall be corrected according to the following equation (4.2) (Bartlett 2001):

$$\text{Corrected SS} = \frac{SS}{(1 + \left(\frac{SS - 1}{Pop}\right))} \quad \text{.......} \quad (4.2)$$

Where Pop = targeted Population (500)

$$\text{Corrected SS} = \frac{43}{(1 + \left(\frac{43 - 1}{500}\right))} = 40$$

The required number of respondents is not less than 40 experts. However, the target number of questionnaire recipients shall consider a percentage of about 50% of no response to the questionnaire, thereafter, the target number of questionnaire recipients will be 40 X (1+50%) = 60 expert.

4.4.4 Questionnaire Distribution and Collection

The questionnaire was randomly distributed among a group of experts (recipients) to resolve, the group resembled principal stakeholders, such as owners, project managers, consultants and contractors for the previously mentioned type of projects. The questionnaire was distributed by email and by hand. A brief explanation to the questionnaire that outlined why it was conducted and instructions on how to complete it was illustrated to the recipients in order to simplify the process. Finally the resolved questionnaires were collected to be analyzed as shown in the following sections.

4.4.5 Questionnaire Response

I. List of Respondents

The step following the collection of the questionnaire respond was making a complete list of respondent, this list includes data such as respondent names, organization name, title/position and years of experience. The complete list of respondent was presented in appendix (A).

II. Classification of the Surveyed Experts (Respondents) Based on Their Experience

The respondents were classified according to their experience, as shown in figure (4.2), the classification indicates that about 25% of the recipients have experience greater than or equals to 25 years, about 32.7% have experience greater than or equals to 20 years, about 9.6% of the recipients have experience greater than or equals to 15 years, about 23.1%

have experience greater than or equals to 10 years, and about 9.6% have experience greater than or equals 5 years.

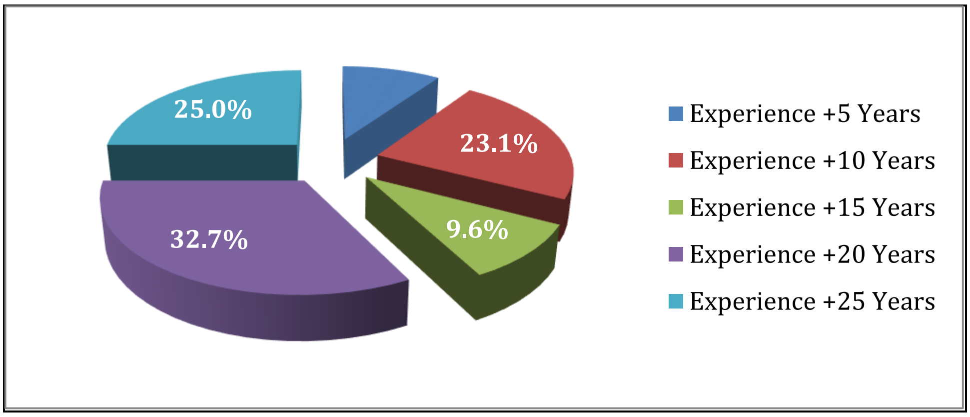


Figure (4.2): Classification of Respondents Based on Their Experience

III. Table of Response

All responses received were then collected in one table; the purpose of the table of responses was to present in one place all the received responses. Table (4.7) represents the responses. A sample of solved questionnaires is presented in Appendix (A)

Table (4.7): Table of Responses

Questionnaire Responses											
No.	Cost Factors	Impact					Degree of Existence				
		V.L	L	M	H	V.H	V.L	L	M	H	V.H
A	General Factors										
1	Project location		1	30	10	11	1	2	35	9	5
2	Site accessibility	1	16	22	10	3		17	27	5	3
3	Site Constraints	1	26	12	10	3	12	17	18	4	1
4	Desired completion time for the project		1	5	36	10		3	10	25	14
5	Subsistence of time constrains			11	35	6		5	16	31	
6	Owner requirements for bid packaging for multiple contractors		15	22	6	9		29	12	11	
7	Environmental impact		28	9	15		2	8	26	15	1

Questionnaire Responses											
No.	**Cost Factors**	**Impact**					**Degree of Existence**				
		V.L	L	M	H	V.H	V.L	L	M	H	V.H
	assurance system requirements										
8	Applying safety system during construction		29	9	7	7		5	9	34	4
B	**Engineering Related Factors**										
1	Site topography		4	28	7	13		6	29	13	4
2	Accumulative built-up area		3	11	11	27		3	14	10	25
3	Buildings closeness (attached, semi-attached or separated)		6	23	19	4		8	16	24	4
4	Other supplementary buildings (W.tank, administration, warehouse,… etc)	1	7	9	34	1		7	13	20	12
5	Accumulative Sterile Areas (total area)	1	5	9	14	23	1	6	13	10	22
6	Structural design loads	1	9	25	11	6		11	18	17	6
7	Geotechnical nature of soil	1	13	19	9	10	1	8	21	17	5
8	Desired structural system		1	21	22	8	1	8	21	16	6
9	Desired HVAC system			11	11	30		5	10	9	28
10	Desired Firefighting system		3	16	7	26		5	14	7	26
C	**Resources Related Factors**										
1	Consultant fees	4	31	9	6	2	2	5	7	30	8
2	Desired level of contractor's prequalification		1	8	36	7		3	15	30	4
3	Contractor overhead		1	17	29	5		3	15	28	6
4	Need for special contractor(s)		1	9	36	6		9	24	16	3
5	Reinforcement price		2	13	20	17		2	14	29	7
6	Cement price		6	32	10	4		3	13	30	6
7	stainless Steel price		17	26	5	4	1	9	14	24	4
8	Labor price		4	32	15	1	1	5	12	30	4
9	% of imported material		2	5	35	10		5	12	29	6
10	Availability of required power		4	11	33	4		3	16	39	4

Questionnaire Responses											
No.	Cost Factors	Impact					Degree of Existence				
		V.L	L	M	H	V.H	V.L	L	M	H	V.H
D	**Special Factors**										
1	Target market (regional or international)		1	5	41	5		7	13	30	2
2	Type of products and type of production method		21	9	16	6		17	10	24	1
3	Special finishing required for sterile areas		1	6	27	18		9	10	5	28
4	Additional requirements for structural system regarding sterile manufacturing		3	18	18	13	1	9	11	8	23
5	Additional requirements for HVAC system regarding sterile manufacturing		1	7	27	17		7	14	7	24
6	Industrial safety requirements (firefighting, fire alarm, .. etc)		2	11	34	5		3	11	24	4
E	**External Factors**										
1	Currency exchange rate			6	14	32		4	11	6	31
2	Inflation		1	8	25	18		6	7	11	28
3	International insurances (if any)		4	14	33	1	3	9	12	5	23

4.4.6 Questionnaire Analysis

The first stage of questionnaire analysis was using severity Index to determine the degree of importance for each factor by calculating the (Importance Index). Thereafter, based on this index the Factors were ranked in an ascending order, facilitating eventually to eliminate the factors with the lowest importance, producing the important Factors.

The second stage of questionnaire analysis was grouping the initial shortlisted factors having logical correlation then selecting one prominent factor from each group as an indicator to produce a final list of Cost Indicators. The analysis took place as follows:

Stage 1

I. Ranking of Cost Factors based on their importance index

The questionnaire respondents provided qualitative scale-based answers to both "The impact" & "The degree of existence" for each factor, stating their opinions in accordance with their experience in the Egyptian construction industrial field specifically industrial projects encompassing sterile buildings, this scale varies from "very low to very high", as indicated in table (4.4). Thereafter, the previous qualitative scale was converted into a quantitative scale varying from 1 to 5 respectively in accordance with the qualitative scale; as illustrated in table (4.5), which represents the qualitative scale and equivalent quantitative value.

The importance of each Cost Factors (Importance Index) was concluded by multiplying the average weighted impact of each factor (Severity Index) times its average weighted degree of existence (Frequency Index) (Hosny 2011), using the following formulas:

- **Severity index:** This index expresses severity of cost factor. It is computed as per equation (4.3)

  $$\mathbf{S.I = \Sigma a * n / N} \quad \text{..........} \quad (4.3)$$

  Where: a = constant expressing the weight assigned to each responses (ranges from 1 for very low to 5 for very high), n = frequency of each response, N = total number of responses.

- **Frequency Index:** This index expresses the degree of existence of cost factors. It is computed as per equation (4.4)

  $$\mathbf{F.I = \Sigma a * n / N} \quad \text{..........} \quad (4.4)$$

  Where: a = constant expressing the weight assigned to each responses (ranges from 1 for very low to 5 for very high), n = frequency of each response, N = total number of responses.

- **Importance Index:** This index expresses the overview of cost factor based on both their degree of existence & severity. This index was calculated as per equation (4.5)

$$\mathbf{IMP.I = F.I * S.I} \quad \ldots\ldots (4.5)$$

Table (4.8) presents the accumulated responses and the calculated severity index, frequency index and the consequently resultant important index.

Table (4.8): Importance Index of Cost Factors

No.	Cost Factors	Impact						Degree of Existence						Importance Index
		V.L	L	M	H	V.H	Severity Index	V.L	L	M	H	V.H	Frequency Index	
		1	2	3	4	5		1	2	3	4	5		
A	**General Factors**													
1	Project location		1	30	10	11	3.60	1	2	35	9	5	3.29	11.83
2	Site accessibility	1	16	22	10	3	2.96		17	27	5	3	2.88	8.54
3	Site Constraints	1	26	12	10	3	2.77	12	17	18	4	1	2.33	6.44
4	Desired completion time for the project		1	5	36	10	4.06		3	10	25	14	3.96	16.07
5	Subsistence of time constrains			11	35	6	3.90		5	16	31		3.50	13.66
6	Owner requirements for bid packaging for multiple contractors		15	22	6	9	3.17		29	12	11		2.65	8.42
7	Environmental impact assurance system requirements		28	9	15		2.75	2	8	26	15	1	3.10	8.51
8	Applying safety system during construction		29	9	7	7	2.85		5	9	34	4	3.71	10.56
B	**Engineering Related Factors**													
1	Site topography		4	28	7	13	3.56		6	29	13	4	3.29	11.70
2	Accumulative built-up area		3	11	11	27	4.19		3	14	10	25	4.10	17.17

No.	Cost Factors	Impact					Severity Index	Degree of Existence					Frequency Index	Importance Index
		V.L	L	M	H	V.H		V.L	L	M	H	V.H		
		1	2	3	4	5		1	2	3	4	5		
3	Buildings closeness (attached, semi-attached or separated)		6	23	19	4	3.40		8	16	24	4	3.46	11.78
4	Other supplementary buildings (W.tank, administration, warehouse,… etc)	1	7	9	34	1	3.52		7	13	20	12	3.71	13.06
5	Accumulative Sterile Areas (total area)	1	5	9	14	23	4.02	1	6	13	10	22	3.88	15.61
6	Structural design loads	1	9	25	11	6	3.23		11	18	17	6	3.35	10.81
7	Geotechnical nature of soil	1	13	19	9	10	3.27	1	8	21	17	5	3.33	10.88
8	Desired structural system		1	21	22	8	3.71	1	8	21	16	6	3.35	12.42
9	Desired HVAC system			11	11	30	4.37		5	10	9	28	4.15	18.13
10	Desired Firefighting system		3	16	7	26	4.08		5	14	7	26	4.04	16.46
C	**Resources Related Factors**													
1	Consultant fees	4	31	9	6	2	2.44	2	5	7	30	8	3.71	9.06
2	Desired level of contractor's prequalification		1	8	36	7	3.94		3	15	30	4	3.67	14.48
3	Contractor overhead		1	17	29	5	3.73		3	15	28	6	3.71	13.85
4	Need for special contractor(s)		1	9	36	6	3.90		9	24	16	3	3.25	12.69
5	Reinforcement price		2	13	20	17	4.00		2	14	29	7	3.79	15.15
6	Cement price		6	32	10	4	3.23		3	13	30	6	3.75	12.12
7	stainless Steel price		17	26	5	4	2.92	1	9	14	24	4	3.40	9.95
8	Labor price		4	32	15	1	3.25	1	5	12	30	4	3.60	11.69
9	% of imported material		2	5	35	10	4.02		5	12	29	6	3.69	14.84
10	Availability of required power		4	11	33	4	3.71		3	16	39	4	4.42	16.42

No.	Cost Factors	Impact						Degree of Existence						Importance Index
		V.L	L	M	H	V.H	Severity Index	V.L	L	M	H	V.H	Frequency Index	
		1	2	3	4	5		1	2	3	4	5		
D	**Special Factors**													
1	Target market (regional or international)		1	5	41	5	3.96		7	13	30	2	3.52	13.94
2	Type of products and type of production method		21	9	16	6	3.13		17	10	24	1	3.17	9.95
3	Special finishing required for sterile areas		1	6	27	18	4.19		9	10	5	28	4.00	16.77
4	Additional requirements for structural system regarding sterile manufacturing		3	18	18	13	3.79	1	9	11	8	23	3.83	14.50
5	Additional requirements for HVAC system regarding sterile manufacturing		1	7	27	17	4.15		7	14	7	24	3.92	16.30
6	Industrial safety requirements (firefighting, fire alarm, .. etc)		2	11	34	5	3.81		3	11	24	4	2.98	11.35
E	**External Factors**													
1	Currency exchange rate			6	14	32	4.50		4	11	6	31	4.23	19.04
2	Inflation		1	8	25	18	4.15		6	7	11	28	4.17	17.33
3	International insurances (if any)		4	14	33	1	3.60	3	9	12	5	23	3.69	13.28

II. Important Factors

The cost factors were then ranked in an ascending order according to their importance index as shown in table (4.9). Furthermore, by examining the importance index of each cost factors, we can see that some factors are heavily considered to have high impact. For example, currency exchange rate, desired HVAC system, inflation, accumulative built-up area, special finishing required for sterile areas and accumulative sterile areas.

However, factors such as need for special contractor, desired structural system and project location, have a medium impact. On the other hand, factors such as applying safety system during construction, consultant fees, owner requirements for bid packaging for multiple contractors, and site constrains, have low impact. Table (4.9) indicates the ranking of all cost factors according to their importance index

Table (4.9): Ranking of Cost Factors

Ranking	Group	Cost Factors	Importance Index
1	E	Currency exchange rate	19.04
2	B	Desired HVAC system	18.13
3	E	Inflation	17.33
4	B	Accumulative built-up area	17.17
5	D	Special finishing required for sterile areas	16.77
6	B	Desired Firefighting system	16.46
7	C	Availability of required power	16.42
8	D	Additional requirements for HVAC system regarding sterile manufacturing	16.30
9	A	Desired completion time for the project	16.07
10	B	Accumulative Sterile Areas (total area)	15.61
11	C	Reinforcement price	15.15
12	C	% of imported material	14.84
13	D	Additional requirements for structural system regarding sterile manufacturing	14.50
14	C	Desired level of contractor's prequalification	14.48
15	D	Target market (regional or international)	13.94
16	C	Contractor overhead	13.85

Ranking	Group	Cost Factors	Importance Index
17	A	Subsistence of time constrains	13.66
18	E	International insurances (if any)	13.28
19	B	Other supplementary buildings (W. tank, administration, warehouse,… etc)	13.06
20	C	Need for special contractor(s)	12.69
21	B	Desired structural system	12.42
22	C	Cement price	12.12
23	A	Project location	11.83
24	B	Buildings closeness (attached, semi-attached or separated)	11.78
25	B	Site topography	11.70
26	C	Labor price	11.69
27	D	Industrial safety requirements (firefighting, fire alarm,. etc)	11.35
28	B	Geotechnical nature of soil	10.88
29	B	Structural design loads	10.81
30	A	Applying safety system during construction	10.56
31	C	stainless Steel price	9.95
32	D	Type of products and type of production method	9.95
33	C	Consultant fees	9.06
34	A	Site accessibility	8.54
35	A	Environmental impact assurance system requirements	8.51
36	A	Owner requirements for bid packaging for multiple contractors	8.42
37	A	Site Constraints	6.44

Thereafter, the relative weight of highest importance index was assumed to be 100%. Then, the relative weight for each cost factor was calculated relatively. Also, all factors have been considered as a ratio of the most important factor. Table (4.10) presents the results of relative weights and ratios calculations.

In principle, weights greater than 60% were considered the highest important factors. Therefore, the top 26 Cost factors and their importance indices were shaded by gray color in table (4.10), the accumulative ratio of the top 26 Cost Factors represent 78.45% of the total value of importance index. These factors were considered as potential input variables for the proposed ANN model that will be presented in the next Chapter. Table (4.10) presents the calculations done to deduce the most important cost factors.

Table (4.10): The Calculations Leading to the Highest Important Factors

No.	Group	Cost Factors	Importance Index	Relative Weights	Ratio	Accumulative Ratio
1	E	Currency exchange rate	19.04	100%	3.93%	
2	B	Desired HVAC system	18.13	95.24%	3.74%	
3	E	Inflation	17.33	91.05%	3.58%	
4	B	Accumulative built-up area	17.17	90.20%	3.54%	
5	D	Special finishing required for sterile areas	16.77	88.08%	3.46%	
6	B	Desired Firefighting system	16.46	86.48%	3.40%	
7	C	Availability of required power	16.42	86.23%	3.39%	
8	D	Additional requirements for HVAC system regarding sterile manufacturing	16.30	85.59%	3.36%	
9	A	Desired completion time for the project	16.07	84.43%	3.32%	
10	B	Accumulative Sterile Areas (total area)	15.61	82.01%	3.22%	
11	C	Reinforcement price	15.15	79.60%	3.13%	

No.	Group	Cost Factors	Importance Index	Relative Weights	Ratio	Accumulative Ratio
12	C	% of imported material	14.84	77.95%	3.06%	78.45%
13	D	Additional requirements for structural system regarding sterile manufacturing	14.50	76.15%	2.99%	
14	C	Desired level of contractor's prequalification	14.48	76.06%	2.99%	
15	D	Target market (regional or international)	13.94	73.23%	2.88%	
16	C	Contractor overhead	13.85	72.73%	2.86%	
17	A	Subsistence of time constrains	13.66	71.77%	2.82%	
18	E	International insurances (if any)	13.28	69.74%	2.74%	
19	B	Other supplementary buildings (W.tank, administration, warehouse,…etc)	13.06	68.61%	2.69%	
20	C	Need for special contractor(s)	12.69	66.64%	2.62%	
21	B	Desired structural system	12.42	65.23%	2.56%	
22	C	Cement price	12.12	63.64%	2.50%	
23	A	Project location	11.83	62.14%	2.44%	
24	B	Buildings closeness (attached, semi-attached or separated)	11.78	61.87%	2.43%	
25	B	Site topography	11.70	61.45%	2.41%	
26	C	Labor price	11.69	61.39%	2.41%	
27	D	Industrial safety requirements (firefighting, fire alarm, .. etc)	11.35	59.62%	2.34%	80.79%
28	B	Geotechnical nature of soil	10.88	57.13%	2.24%	19.21%
29	B	Structural design loads	10.81	56.78%	2.23%	
30	A	Applying safety system during construction	10.56	55.49%	2.18%	
31	C	stainless Steel price	9.95	52.26%	2.05%	
32	D	Type of products and type of production method	9.95	52.24%	2.05%	

No.	Group	Cost Factors	Importance Index	Relative Weights	Ratio	Accumulative Ratio
33	C	Consultant fees	9.06	47.61%	1.87%	
34	A	Site accessibility	8.54	44.87%	1.76%	
35	A	Environmental impact assurance system requirements	8.51	44.72%	1.76%	
36	A	Owner requirements for bid packaging for multiple contractors	8.42	44.23%	1.74%	
37	A	Site Constraints	6.44	33.85%	1.33%	
Total			484.78		100.00 %	100.00%

In addition to the previous consideration, Pareto's principle "The 80-20 rule" was also applied to determine the important cost factors. Consequently, the important factors will considered to be the group of factors which produce accumulative Ratio equal to 80%. figure (4.3) illustrates graphically the application of 80-20 rule, it shows that at ratio 80% resulted at factor 27. Therefore, beside the first 26 factors deduced previously and shaded in gray, the factor No. (27) will also be considered as an important factor, this factor was shaded in green in table (4.10).

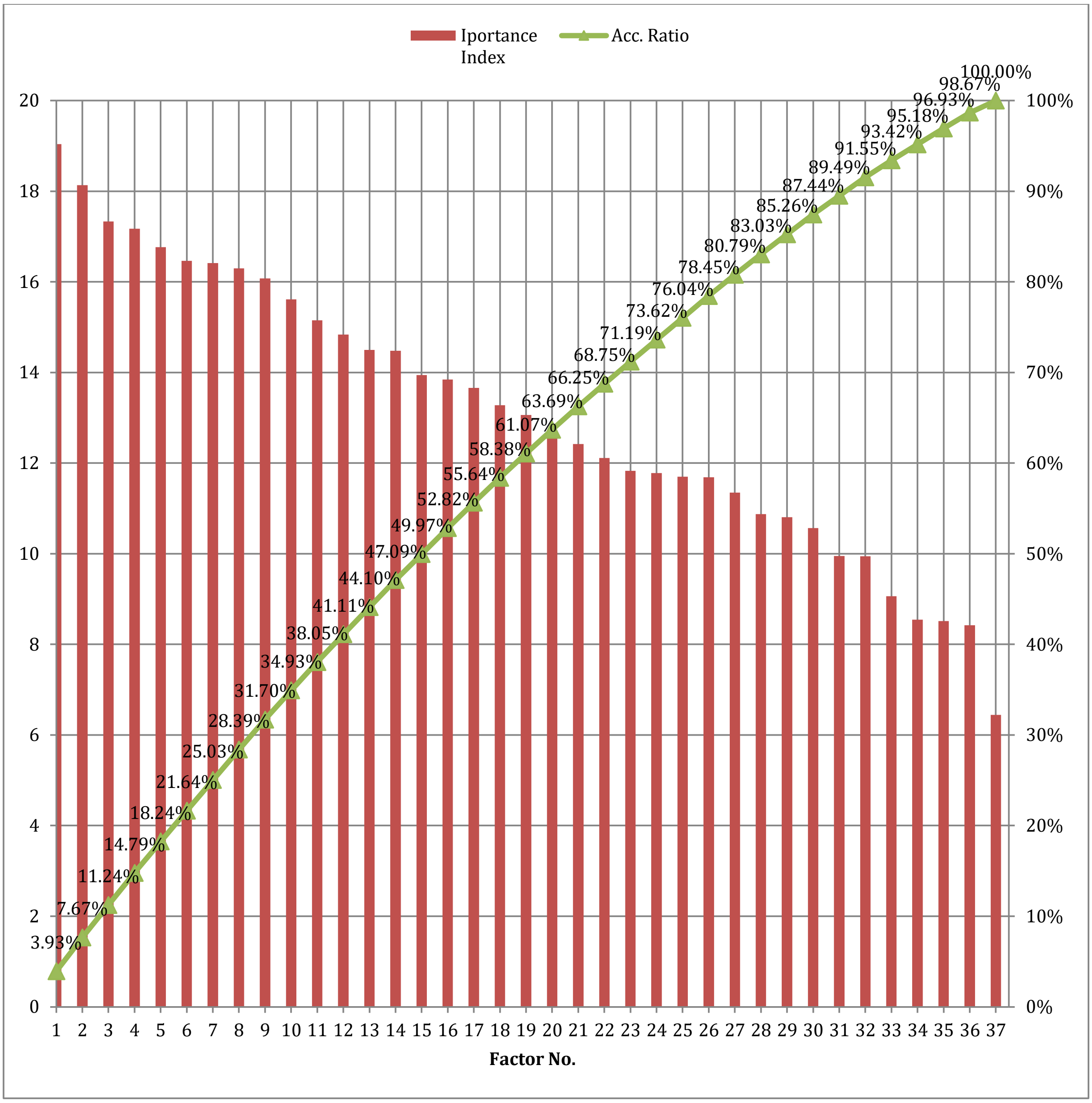


Figure (4.3): Pareto's Principle "The 80-20 Rule"

The list of deduced factors (27 factor) which considered as most important cost factors were presented separately in table (4.11). The main focus will be directed to these factors in the next stage of analysis, and then the remaining factors will be discussed consequently.

Table (4.11): The Highest (Most) Important Factors

No.	Group	Cost Factors
1	E	Currency Exchange rate
2	B	Desired HVAC system
3	E	Inflation
4	B	Accumulative built-up area
5	D	Special finishing required for sterile areas
6	B	Desired Firefighting system
7	C	Availability of required power
8	D	Additional requirements for HVAC system regarding sterile manufacturing
9	A	Desired completion time for the project
10	B	Accumulative Sterile Areas (total area)
11	C	Reinforcement price
12	C	% of imported material
13	D	Additional requirements for structural system regarding sterile manufacturing
14	C	Desired level of contractor's prequalification
15	D	Target market (regional or international)
16	C	Contractor overhead
17	A	Subsistence of time constrains
18	E	International insurances (if any)
19	B	Other supplementary buildings (W. tank, administration, warehouse,… etc)
20	C	Need for special contractor(s)
21	B	Desired structural system
22	C	Cement price
23	A	Project location
24	B	Buildings closeness (attached, semi-attached or separated)
25	B	Site topography
26	C	Labor price
27	D	Industrial safety requirements (fire fighting, fire alarm, .. etc)

Stage 2

I. Selection of the Cost Indicators

An extensive study was conducted on through the most important cost factors (concluded from Stage 1). This study aims to group factors having logical correlation and to select one prominent factor from each group as an indicator. Accordingly, the final list of cost indicators will be deduced. The main purpose of this process is to eliminate redundancy and simplify the input data required for the creation of the Neural Model.

This study deduced that, the consumer price index can be an indicator for reinforcement price and other material prices, labor prices and inflation, due to the proportional relation between these factors.

Also, the accumulative sterile areas directly influence the amount of special finishing required for sterile areas, additional requirements for HVAC system regarding sterile manufacturing and additional requirements for structural system regarding sterile manufacturing, due to the same relationship between these cost factors and the selected cost indicator.

In addition, project location found to be a realistic cost indicator representing Site Topography and Availability of Required Power, also the Target Market strongly affects the selection of Desired HVAC system, % OF Imported materials and the Desired Fire Fighting system.

Furthermore, contractor overhead and need for special contractor can be indicated by desired level of contractor's prequalification according to the inevitable logical relation between the contractor overhead and the contractor prequalification.

Finally, desired completion time for the project was selected to represent the subsistence of time constrain.

In addition to the concluded Cost Indicators, a new Cost Factor (Project Status) emerged as a result of multiple suggestions from respondents. This cost factor was found to be of a great value to the Neural Model, as it indicates the status of the project as a new project, extension or renovation. This factor was presented in table (4.12) as a cost indicator No. (13).

II. Other Remaining Factors

Moreover, by studying the remaining cost factors it was ratiocinated that; Consumer Price Index is also indicates stainless Steel price and Consultant fees. While, Site Constraints, Site accessibility, Geotechnical nature of soil can be indicated by Project Location which removes any risk factors inherent in unknown project location. Also, Desired completion time affects the owner's selection of bidding type, as well as, the Type of products and type of production method influences the accumulative built-up area.

Furthermore, the Structural design loads is a common attribute in all industrial projects for both production and warehousing areas. Finally, any industrial projects must fulfill the authorities' regulations regarding the Environmental requirements, whilst, Applying safety system during construction is mandatory in all food and pharmaceutical company in order to keep its reputation at local and of course the international market.

As a conclusion, the previous discussion proved that the selected cost indicators are capable to present both the group of factors with highest impotence and the other remaining factors. Accordingly, they can be used in estimating process. Table (4.12) represents the selected Cost Indicators for all important cost factors.

Table (4.12): The Cost Indicators

No.	Cost Indicator	Grouped Factors (Factors Having a Logical Correlation)
1	Currency Exchange Rate	• Currency exchange rate
2	Consumer Price Index	• Reinforcement price
		• Cement price
		• Labor price
		• Inflation
3	Accumulative Built-up Area	• Accumulative built-up area
4	Accumulative Sterile Areas (Total Area)	• Accumulative Sterile Areas (total area)
		• Special finishing required for sterile areas
		• Additional requirements for HVAC system regarding sterile manufacturing
		• Additional requirements for structural system regarding sterile manufacturing
5	Project Location	• Project location
		• Site topography
		• Availability of required power
6	Target market (Regional or international)	• Target market (regional or international)
		• Desired HVAC system
		• Desired Firefighting system
		• % of imported material
7	International Insurances (if any)	• International insurances (if any)
8	Desired Level of Contractor's Prequalification	• Desired level of contractor's prequalification
		• Contractor overhead
		• Need for special contractor(s)
9	Desired Completion Time for the Project	• Desired completion time for the project
		• Subsistence of time constrains
10	Other Supplementary Buildings (W. Tank, Administration, Warehouse,… etc)	• Other supplementary buildings (W. tank, administration, warehouse,… etc)
11	Desired structural system	• Desired structural system
12	Buildings closeness (attached, semi-attached or separated)	• Buildings closeness (attached, semi-attached or separated)
13	Project Status	• Project Status

These cost indicators will be used in two crucial actions; first one is historical data collection process, the collected projects must full fill those indicators, otherwise, the project will be neglected. Second on is developing the proposed ANN model; these indicators will be the input data for the model.

4.5 Historical Data Collection

Based on thesis methodology and after determining the most important factors, data for training and testing the proposed artificial neural networks model were collected. These data were gathered from (18) real life projects conducted from 1999 to 2013 in Egypt. This process was done using a data collection form as shown in table (4.13).

Table (4.13): The Data Collection Form

		Company Name				
		Project Number				
I	**General Data**					
	1	Project Name				
	2	Project Duration				
	3	Start Date		Finish Date		
	4	Project Location				
	5	Total Project Cost				
	6	Total Number of buildings				
	7	Buildings Names				
	8	Buildings Closeness	☐Attached	☐Smi-Attached	☐Separated	
	9	Industrial Field	☐Medical	☐Food		
	10	Project Status	☐New	☐Extension	☐Renovation	
	11	Target Market	☐Regional	☐International		
	12	Exchange Rate				
	13	Average Resources Rates	Reinforcement (ton)		Concrete (M3)	
			stainless St. (ton)			
II	**Buildings Description**					
	a	Building Number ( )				
	1	Building Name / Use				
	2	Repetition	☐No	☐Yes	Number	
	3	Structural system	☐R.C	☐Steel	☐Mixed	
	4	Total Built Area		Area peer Floor		
	5	Additional Structural requirements	☐No	☐Yes		
	6	Sterile Area	☐No	☐Yes	Area	
	7	Additional Architectural requirements	☐No	☐Yes		
	8	Air Conditioning	☐No	☐Yes		
	9	Fire Fighting	☐No	☐Yes		
	10	Required Power Availability	☐Not Available	☐Available	☐ Smi-Available	
	11	Additional Electro-Mechanical requirements	☐No	☐Yes		

This form was used to collect the data of industrial projects, especially pharmaceutical and food type of projects that comprising sterile buildings. These projects were designed and constructed in Egypt. The data collection process was done based on the required cost indicators which presented in table (4.13). A schematic description for the buildings included in the targeted projects ware presented in appendix (B).

In this study, only projects with complete and reliable data were considered to be an input fact for the proposed ANN model. Other projects with incomplete or doubtful data were completely neglected. This constrain explains the scarcity of accepted projects, accordingly, only (18) real projects were selected to be part of this study. Table (4.14) shows the list of accepted projects. The full collected data for these projects were presented in appendix (B).

Table (4.14): The List of Accepted Projects

No.	Project Name	Description	Start Date	Location	Project Type
1	Global Nabi Factory	New Project	9-2007	6th of October- Giza	Pharmaceutical
2	Nestle Ice Cream Factory	Extension	5-2012	6th of October-Giza	Food
3	Novartis Pharma factory	Renovation	4-2006	Sawah-Cairo	Pharmaceutical
4	Nestle Dry Goods factory	New Project	5-2013	6th of October-Giza	Food
5	Sakara factory	New project	2-2011	Sadat City	Food
6	Juhayna Egyfood factory	Extension	2-2012	6th of October-Giza	Food
7	Cadboury Diego factory	New project	3-2009	10th of Ramadan City	Food
8	Juhayna Concentrates factory	New project	9-2007	6th of October-Giza	Food
9	Nestle Nesquick factory	Extension	2-2009	6th of October-Giza	Food

No.	Project Name	Description	Start Date	Location	Project Type
10	Nestle Cereals factor	Extension	4-2013	6^{th} of October-Giza	Food
11	Juhayna Moder food factory	New project	3-2007	6^{th} of October-Giza	Food
12	T3A factory	New project	1-1999	Assiut City	Pharmaceutical
13	Servier Packaging building	New project	7-2003	6^{th} of October-Giza	Pharmaceutical
14	Servier factory	Renovation	1-2007	6^{th} of October-Giza	Pharmaceutical
15	Eli-Lilly factory warehouse	New project	3-2009	6^{th} of October-Giza	Pharmaceutical
16	Glaxo Smithkline laboratory	New project	2-2007	Salam City-Cairo	Pharmaceutical
17	Nestle Ice Cream Biscuit factory	New project	11-2012	6^{th} of October-Giza	Food
18	Acdima Pharma	New project	1-2013	6^{th} of October-Giza	Pharmaceutical

Moreover, during this study another two projects were completed with a full data archiving, these two projects were used separately to validate the proposed model. Table (4.15) represents the new projects.

Table (4.15): The New Projects

No.	Project Name	Description	Start Date	Location	Project Type
1	Nerhadou Factory	New Project	10-2013	6^{th} of October-Giza	Pharmaceutical
2	Doehler Factory	New Project	11-2013	6^{th} of October-Giza	Food

For each projected, thirteen input criterion from I1 to I13 were tabulated as a cost indicator to be used in developing the proposed ANN model. Table (4.16) represents the assigned input character for each cost indicator. Moreover, the full data for all projects were concluded in appendix (B).

Table (4.16): The Assigned Input Characters

No.	Cost Indicator	Assigned Input Character
1	Currency Exchange Rate	I_1
2	Consumer Price Index	I_2
3	Desired Completion Time for the Project	I_3
4	Accumulative Built-up Area	I_4
5	Accumulative Sterile Areas (total area)	I_5
6	Other Supplementary Buildings (W. Tank, Administration, Warehouse,… etc)	I_6
7	Desired Structural System	I_7
8	Buildings Closeness (attached, semi-attached or separated)	I_8
9	Project Status	I_9
10	Project Location	I_{10}
11	Target Market (Regional or International)	I_{11}
12	International Insurances (if any)	I_{12}
13	Desired Level of Contractor's Prequalification	I_{13}

4.6 Conclusion

Heretofore, it has been pointed out that the list of factors affecting total construction cost which concluded in Chapter (2) and the additional factors derived from interviews were ranked according to its importance index. Then, the correlation between these factors logically studied to group all related factors under on cost indicator. Thereafter, thirteen cost indicators were concluded out of (37) cost factors. These indicators were considered as the most important factors affecting final construction cost.

Also, data were collected from (18) real life projects conducted in Egypt at the field of industrial projects. All projects with incomplete or doubtful data were neglected from this study Extremeness. Finally, this Chapter provides the cost indicators and the historical facts required for the development of the proposed cost neural network model which is the issue of the next Chapter.

CHAPTER 5

DEVELOPMENT OF ARTIFICIAL NEURAL NETWORK MODEL

CHAPTER 5

DEVELOPMENT OF ARTIFICIAL NEURAL NETWORK MODEL

5.1 Introduction

As mentioned in the introductory Chapter, the fifth chapter is parametric cost estimating model development using artificial neural networks. This Chapter was prepared to elucidate the research approach in using the technique of the neural network model in predicting the expected project's total cost.

The first step in this chapter was designing the ANN model by using Microsoft Excel as data base software and Evolver add-in as Genetic Algorithms optimizer. Moreover, trial and error practices were implemented to select the best structure of the model. Thereafter, the developed models were trained to select the best one. Finally, the validity of the best model will be tested.

5.2 Design of ANN Model

Design of the neural network model includes two main steps, the first one is selecting the used Software(s) for both modeling and simulation (optimizing) and the second one is determining the network architecture, the steps are as follows:

5.2.1 Selection of Used Software(s)

For the modeling Microsoft Excel 2010 was the selected as data-base software to develop the proposed neural network model; this software is running under Microsoft windows 7 operating system. Microsoft Excel is designed to be user friendly; allowing its users to simply construct a neural network model without having extensive programming knowledge. However, it is not designed for novice users.

For the optimization process, Evolver add-in version 5.1.1 was added to Microsoft Excel, this software produced by Palisade Corporation. Evolver uses Genetic Algorithms as a technique to search for the near optimal solution throughout the optimization process.

One of the main advantages of this software is simplicity of usage for its user; it easily allows the user to design and apply any required constrains with any number of constrains, also Evolver allows its users to customize the values of some important features such as the permutation and crossover, finally the installation compatibility of this software with all Microsoft Excel versions is a vital attribute.

5.2.2 Neural Network Architecture

The structure of the neural network was designed to consist of three parts; the input layer, one or two hidden layers and the output layer, each layer encloses a certain number of nodes, each node in the input layer and output layer were linked to all nodes enclosed in the hidden layers by a different weight.

The Input Layer, the first one, enclosing 13 nodes, each node represents one of the selected cost indicators, as concluded in the previous chapter, each cost indicator will be determined by a value; for quantitative indicators, the exact value will be used, however, for qualitative indicators, a preselected value will be used as an indicator for each case, table (5.1) illustrates the cost indictors used in the input layer and the correspondent determination value for each cost indicator. Finally all the data in this layer will be scaled from (-1) to (1).

Table (5.1): The Input Layer and the Determination Value for Each Cost Indicator

No.	Code	Cost Indicator	Determination Value
1	I_1	Currency exchange rate	Exact Value
2	I_2	Consumer price index	Exact Value
3	I_3	Desired completion time for the project	Exact Value
4	I_4	Accumulative built-up area	Exact Value
5	I_5	Accumulative Sterile Areas (total area)	Exact Value
6	I_6	Other supplementary buildings (W. tank, administration, warehouse,… etc) (total area)	Exact Value
7	I_7	Desired structural system	(1) For mixed (2) For concrete (3) For Steel
8	I_8	Buildings closeness (attached, semi-attached or separated)	(1) For attached (2) For semi-attached (3) For separated
9	I_9	Project Status	(1) For renovation (2) For extension (3) For new building in existing project (4) For entirely new project
10	I_{10}	Project location	(1) For inside Cairo (2) For areas outside Cairo and till 10th of Ramadan city (3) For areas farther than 10th of Ramadan city
11	I_{11}	Target market (regional or international)	(1) For regional market (2) For international market
12	I_{12}	International insurances (if any)	(1) No (2) Yes
13	I_{13}	Desired level of contractor's prequalification	(1) Normal (2) Moderate (3) High

The Hidden Layer(s), the second ordered layer. In this layer(s), the number on nodes (hidden nodes) ware calculated by considering one guidance that the number of hidden nodes must be not less than half the summation of the number of nodes in the input and output layers (Hosny 2011). According to that guidance, the number of hidden nodes will be calculated according to equation (5.1).

$$Min. No. of hidden nodes = No. of input nodes + No. of output nodes \quad (5.1)$$

$$Min No. of hidden nodes = (13 + 1) \div 2 = 7$$

According to the previous guidance, 8 hidden nodes ware used in this layer, also an activation function will be used to activate data derived into these 8 hidden nodes. In the trial and error practices, another hidden layer will be added to a new model to be used in a deferent set of trials, the number of hidden nodes in this layer will not be strict to the above mentioned guidance; only 4 hidden nodes will be used in this layer.

The Output Layer, the third layer, this layer encloses only one output neuron representing “Predicted Cost”, considering the scaling done in the input layer, data in this layer will be scaled back.

5.2.3 Neural Network Type

This thesis tends to use feed forward artificial neural network with objective function. Based on that, a supervised learning technique will be used; where Inputs were fed to the proposed network model and the outputs then calculated. The differences between the calculated outputs and the actual outputs were then evaluated until the learning rule is attained. Learning rule is a procedure for modifying the weights of the network to produce a desirable target.

5.3 Modeling Data

Based on the historical data illustrated in chapter (4) and attached in appendix (B) along with the designed Artificial Neural Network, previously illustrated in item (5.2.) herein above, a set of steps for Data Modeling were carried out to develop the required model, these steps are as follows:

5.3.1 Data Categorization

Historical data for 18 individual projects with the specific type and nature targeted by this study were previously collected for this study, as illustrated in chapter (4) and attached in appendix (B), these historical data were sorted into two categories, training and testing data. First category, the training set of projects' data, represents about 75% - 80% of all collected data, accordingly, 14 projects' data were randomly selected to be the training set of data, these set of data will be used to train the model by reducing the difference between the actual cost and the predicted cost by calculating then minimizing the Root of Mean square Error (RMSE), also the percentage error for this group will be calculated to monitor the average percentage error results during training process.

Consequently, the second category, the testing set of projects' data, will be 4 projects' data representing 20% - 25% of collected data, this set will be used to monitor the results produced during training the model and the average percentage error for this group will be calculated to validate the victor model.

5.3.2 Model Execution

1. Data Input

All projects' data were inserted into Microsoft Office Excel in a table consisting of fourteen horizontal input fields for each project; one for the actual cost and thirteen fields representing the cost indicators. Figure (5.1) shows the input data.

Step 1 :Original unscaled inputs

Set	No.	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	Actual Cost
TR	1	5.50	73.70	12	9200	3450	1100	1	2	4	2	1	1	2	70,000,000.00
TR	2	6.03	123.80	8	6800	0	400	3	1	4	2	2	1	3	67,000,000.00
TR	3	5.70	63.60	8	1200	300	0	2	1	1	1	2	1	2	6,000,000.00
TR	4	6.90	134.80	6	2600	0	0	2	1	3	2	2	1	2	15,000,000.00
TR	5	5.90	110.90	10	1400	0	490	3	3	4	3	2	1	2	27,000,000.00
TR	6	6.01	121.10	18	15000	1000	7850	2	2	3	2	2	1	2	330,000,000.00
TR	7	5.60	90.80	6	3800	0	0	2	1	3	2	2	1	2	25,000,000.00
TR	8	5.50	73.70	5	1900	0	850	3	3	4	2	1	1	2	35,000,000.00
TR	9	5.55	89.70	5	300	300	0	2	2	2	2	2	1	2	6,000,000.00
TR	10	6.75	133.80	4	1000	0	0	2	2	2	2	2	1	2	8,000,000.00
TR	11	5.60	70.80	24	24900	0	6000	2	2	4	2	2	1	2	130,000,000.00
TR	12	3.40	46.40	16	8230	4000	3015	2	3	4	3	2	2	2	43,000,000.00
TR	13	5.67	70.20	8	300	300	20	2	1	1	2	2	2	2	7,000,000.00
TR	14	6.01	52.40	8	1600	0	275	2	2	2	2	2	2	2	16,000,000.00
TS	15	5.60	90.80	7	1900	140	0	2	3	3	2	2	2	2	35,000,000.00
TS	16	5.60	70.50	6	1700	120	0	2	1	2	1	2	2	2	11,000,000.00
TS	17	6.10	125.60	5	400	0	0	3	3	3	2	2	2	1	5,000,000.00
TS	18	6.40	127.90	15	10800	500	5600	1	3	4	2	2	2	2	80,000,000.00

Figure (5.1): Input Data

2. Data Scaling

All input data ware scaled to a range from (-1 to 1) using maximum and minimum values of each input filed to suit Neural Networks processing. This was done by constructing a second table using a linear equation (5.2) for scaling. Figure (5.2) shows the scaled data.

$$\text{Scaled Value} = \frac{2*(\text{Original value} - \text{Min.value})}{(\text{Max.value} - \text{Min.value})} - 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (5.2)$$

Step 2 :Scaled inputs

Set	No.	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	Actual Cost
TR	1	0.2	-0.38235	-0.2	-0.27642	0.725	-0.71975	-1	0	1	0	-1	-1	0	-0.6
TR	2	0.502857	0.751131	-0.6	-0.47154	-1	-0.89809	1	-1	1	0	1	-1	1	-0.618461538
TR	3	0.314286	-0.61086	-0.6	-0.92683	-0.85	-1	0	-1	-1	-1	1	-1	0	-0.993846154
TR	4	1	1	-0.8	-0.81301	-1	-1	0	-1	0.3333	0	1	-1	0	-0.938461538
TR	5	0.428571	0.459276	-0.4	-0.91057	-1	-0.87516	1	1	1	1	1	-1	0	-0.864615385
TR	6	0.491429	0.690045	0.4	0.195122	-0.5	1	0	0	0.3333	0	1	-1	0	1
TR	7	0.257143	0.004525	-0.8	-0.71545	-1	-1	0	-1	0.3333	0	1	-1	0	-0.876923077
TR	8	0.2	-0.38235	-0.9	-0.86992	-1	-0.78344	1	1	1	0	-1	-1	0	-0.815384615
TR	9	0.228571	-0.02036	-0.9	-1	-0.85	-1	0	0	-0.333	0	1	-1	0	-0.993846154
TR	10	0.914286	0.977376	-1	-0.94309	-1	-1	0	0	-0.333	0	1	-1	0	-0.981538462
TR	11	0.257143	-0.44796	1	1	-1	0.528662	0	0	1	0	1	-1	0	-0.230769231
TR	12	-1	-1	0.2	-0.35528	1	-0.23185	0	1	1	1	1	1	0	-0.766153846
TR	13	0.297143	-0.46154	-0.6	-1	-0.85	-0.9949	0	-1	-1	0	1	1	0	-0.987692308
TR	14	0.491429	-0.86425	-0.6	-0.89431	-1	-0.92994	0	0	-0.333	0	1	1	0	-0.932307692
TS	15	0.257143	0.004525	-0.7	-0.86992	-0.93	-1	0	1	0.3333	0	1	1	0	-0.815384615
TS	16	0.257143	-0.45475	-0.8	-0.88618	-0.94	-1	0	-1	-0.333	-1	1	1	0	-0.963076923
TS	17	0.542857	0.791855	-0.9	-0.99187	-1	-1	1	1	0.3333	0	1	1	-1	-1
TS	18	0.714286	0.843891	0.1	-0.14634	-0.75	0.426752	-1	1	1	0	1	1	0	-0.538461538

Figure (5.2): Scaled Data

3. Weighted Input Sum

Each input was connected to all hidden nodes by a weight. All weights were inserted into a table and assigned to be equal to 1 in the beginning of the modeling. This table was assigned to be adjustable cells in the Evolver. In the model with two hidden layers, each node in the first layer was also connected to all nodes in the second layer using the same previous concept. The summation of weighted inputs was then calculated using equation (5.3). Figure (5.3) shows the weights of hidden nodes and the weighted input summation.

$$X_h = \sum_{i=1}^{13} (\mathrm{I}_i * W_{hi}) \quad \text{.......... (5.3)}$$

Step 3.1 :Weights of 8 hidden nodes

Nodes	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13
H1.1	100	-100	100	100	-100	100	-53.8201	75.103	-100	-100	-100	-100	100
H1.2	0.9875	1.32066	1.014027	1.0048	0.930495	1.047089	0.94866	0.9374	1.0753	1.0826	1.0208	1.1071	-8.054
H1.3	100	-95.4407	99.99985	100	-96.7862	-100	100	10.746	35.949	42.859	-100	100	-100
H1.4	-100	100	99.99973	91.126	-98.1157	-100	-16.2705	100	-25.36	99.999	100	-0.393	100
H1.5	1.0356	28.98178	1.631235	100	10.45432	10.9851	-0.44891	8.7793	-5.668	1.5457	-0.151	-1.34	88.567
H1.6	-40.05	-15.4763	100	-100	-1.55991	100	99.91598	-13.87	-100	78.836	21.084	7.6077	-100
H1.7	100	-7.92515	-2.82418	-100	-31.1965	16.84896	99.99985	98.635	100	70.178	38.991	100	100
H1.8	1.0033	0.974055	1.020828	1.0213	1.018512	0.873769	0.975057	1.0042	0.9889	0.9753	1.0065	1.0012	1.0647

Step 3.2 :Calculation of weighted input summation

Set	No.	H1.1	H1.2	H1.3	H1.4	H1.5	H1.6	H1.7	H1.8
TR	1	19.93601	-2.86824	-53.4	-211.281	-42.8979	-296.163	-122.5	-2.543
TR	2	-250.71	-8.51837	-116.8	168.5432	28.69654	-205.597	249.6	0.3685
TR	3	49.73016	-7.05567	-170.2	-127.819	-134.379	-20.6172	-189.5	-6.541
TR	4	-269.734	-1.96279	-158.7	35.96896	-83.5075	-158.655	63.672	-2.231
TR	5	-300.356	1.819977	41.823	324.4861	-92.6262	19.18869	455.66	1.7046
TR	6	156.3204	2.852239	-196.8	118.6389	45.72989	71.05003	27.804	2.4722
TR	7	-234.716	-3.91302	-128.2	19.59747	-103.371	-123.252	-12.48	-3.846
TR	8	24.17976	-3.0115	201.33	-92.2859	-114.241	-124.53	290.2	-2.722
TR	9	-146.77	-4.00137	-194.9	86.22618	-118.614	-50.7046	40.887	-3.794
TR	10	-167.282	-2.19034	-211.4	127.3319	-85.0281	-109.067	100.82	-2.331
TR	11	323.3804	2.293224	148.34	240.8954	79.7885	-35.4612	5.5357	2.3013
TR	12	-463.606	3.448548	-4.135	186.938	-54.1468	79.97179	315.59	3.656
TR	13	-273.723	-3.64682	48.827	-19.1367	-138.465	79.63668	85.176	-3.506
TR	14	-173.52	-2.29788	159.99	22.1073	-135.22	-6.28541	268.3	-2.028
TS	15	-296.957	0.187269	81.031	207.8669	-103.045	-110.437	397.77	0.1799
TS	16	-245.196	-4.22053	25.889	-131.653	-133.575	-96.1197	69.515	-4.039
TS	17	-526.134	10.12154	209.04	117.5133	-182.201	58.15503	435.05	0.7438
TS	18	-70.9931	4.715144	-37.14	231.0524	9.641879	-158.281	347.31	4.1259

Figure (5.3): Weighted Input Summation

4. Activation of Weighted Input Sum

The activation process was created based on changeable cell for activation function, the activation function can be selected manually from group of preset functions, this group of functions includes 'Tanh', 'Sigmoid' and 'linear [F(x) =X+0]' activation functions. Consequently the previously calculated input summations were activated three times separately according to the trial and error practices using equation (5.4).Thereafter, the model results for each trial were saved in order to select the best fit activation function. The outputs from this step “activated summation” are considered inputs for the next step; either for the second hidden layer or the output layer. In case of second hidden layer, the activated summations of this layer are the inputs for the Output layer. Figure (5.4) presents the activated summation.

$$H_{h\ activated} = f(X_h) \quad (5.4)$$

Step4 :Activation of weighted input summation (Output of hidden nodes)

		Activation Fun	Non						
Set	No.	H1	H2	H3	H4	H5	H6	H7	H8
TR	1	19.936	-2.86824	-53.4	-211.281	-42.8979	-296.163	-122.5	-2.543
TR	2	-250.71	-8.51837	-116.8	168.543	28.6965	-205.597	249.6	0.369
TR	3	49.7302	-7.05567	-170.2	-127.819	-134.379	-20.6172	-189.5	-6.541
TR	4	-269.734	-1.96279	-158.7	35.969	-83.5075	-158.655	63.67	-2.231
TR	5	-300.356	1.81998	41.82	324.486	-92.6262	19.1887	455.7	1.705
TR	6	156.32	2.85224	-196.8	118.639	45.7299	71.05	27.8	2.472
TR	7	-234.716	-3.91302	-128.2	19.5975	-103.371	-123.252	-12.48	-3.846
TR	8	24.1798	-3.0115	201.3	-92.2859	-114.241	-124.53	290.2	-2.722
TR	9	-146.77	-4.00137	-194.9	86.2262	-118.614	-50.7046	40.89	-3.794
TR	10	-167.282	-2.19034	-211.4	127.332	-85.0281	-109.067	100.8	-2.331
TR	11	323.38	2.29322	148.3	240.895	79.7885	-35.4612	5.536	2.301
TR	12	-463.606	3.44855	-4.135	186.938	-54.1468	79.9718	315.6	3.656
TR	13	-273.723	-3.64682	48.83	-19.1367	-138.465	79.6367	85.18	-3.506
TR	14	-173.52	-2.29788	160	22.1073	-135.22	-6.28541	268.3	-2.028
TS	15	-296.957	0.18727	81.03	207.867	-103.045	-110.437	397.8	0.18
TS	16	-245.196	-4.22053	25.89	-131.653	-133.575	-96.1197	69.52	-4.039
TS	17	-526.134	10.1215	209	117.513	-182.201	58.155	435	0.744
TS	18	-70.9931	4.71514	-37.14	231.052	9.64188	-158.281	347.3	4.126

TANH
SIG
Non

Figure (5.4): Activated Summation

5. Weighted Output Sum and Activation

The hidden nodes, in the first or second hidden layer, where then connected to the Output node by weights, all weights were inserted into a table and assigned to be equal to 1 in the beginning of the modeling, this table was assigned to be a adjustable cells in the Evolver, the weighted summation were then calculated to be the input for the output node using equation (5.5). Thereafter an exit activation function was used to activate the weighted summations, as demonstrated in the previous step no.(4), the weighted summation where activated by selecting one of the preset activation functions using equation (5.6). The output from this step is the scaled predicted cost. Figure (5.5) represents the activated output summation.

$$X_o = \sum_{h=1}^{N}(H_h * W_{ho}) \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \;(5.5)$$

$$O \;= f(X_o) \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \;(5.6)$$

Step5.1 :Output weights (from 8 hidden nodes to one output)

Nodes	H1	H2	H3	H4	H5	H6	H7	H8
Wo	-1.269753	-0.681033973	-0.572024662	1.811304014	-0.893292933	1.571381013	0.717120472	-1.198630383

Step 5.1 :Weighted Summation & Its Activation

Step6 :Output

Activation Function				Non
Set	No.	Weighted sum	Output	
TR	1	-0.603853729	-0.603853729	
TR	2	-0.613574318	-0.613574318	
TR	3	-0.994778961	-0.994778961	
TR	4	-0.952895186	-0.952895186	
TR	5	-0.865674691	-0.865674691	
TR	6	0.992648967	0.992648967	
TR	7	-0.850987501	-0.850987501	
TR	8	-0.814135297	-0.814135297	
TR	9	-1.032319137	-1.032319137	
TR	10	-0.968214588	-0.968214588	
TR	11	-0.222203942	-0.222203942	
TR	12	-0.763490832	-0.763490832	
TR	13	-0.985068061	-0.985068061	
TR	14	-0.937118784	-0.937118784	
TS	15	-0.049667777	-0.049667777	
TS	16	-0.964642441	-0.964642441	
TS	17	-0.999370992	-0.999370992	
TS	18	-0.534825581	-0.534825581	80590843.16

Figure (5.5): Activated Output Summation

6. Output

The previously calculated outputs "Scaled Predicted Cost" were then interpreted by scaling back calculation, using equation (5.7). This operation was done to deduce the predicted cost; which is the target of this thesis. Figure (5.6) represents the scaled back output "Predicted Cost".

$$\text{Scaled Value} = \frac{(\text{output value}+1)*(\text{Max.value}-\text{Min.value})}{2} + \text{Min. value} \ldots\ldots\ldots(5.7)$$

				Step 6 :Output
		Activation Function		Non
Set	No.	/eighted su	Output	Scaled back output (P. cost)
TR	1	-0.6038537	-0.6038537	69,373,769.10
TR	2	-0.6135743	-0.6135743	67,794,173.38
TR	3	-0.994779	-0.994779	5,848,418.77
TR	4	-0.9528952	-0.9528952	12,654,532.32
TR	5	-0.8656747	-0.8656747	26,827,862.65
TR	6	0.99264897	0.99264897	328,805,457.09
TR	7	-0.8509875	-0.8509875	29,214,531.03
TR	8	-0.8141353	-0.8141353	35,203,014.23
TR	9	-1.0323191	-1.0323191	-251,859.84
TR	10	-0.9682146	-0.9682146	10,165,129.41
TR	11	-0.2222039	-0.2222039	131,391,859.36
TR	12	-0.7634908	-0.7634908	43,432,739.73
TR	13	-0.9850681	-0.9850681	7,426,440.03
TR	14	-0.9371188	-0.9371188	15,218,197.55
TS	15	-0.0496678	-0.0496678	159,428,986.30
TS	16	-0.9646424	-0.9646424	10,745,603.30
TS	17	-0.999371	-0.999371	5,102,213.87
TS	18	-0.5348256	-0.5348256	80,590,843.16

Figure (5.6): Predicted Cost

7. Calculating Error

The difference between the Predicted cost and the Actual cost (Error) was calculated for the training set, thereafter the Root of Mean Square Error (RMSE) was calculated according to equation (5.8).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Pi-Ai)^2}{n}} \quad \text{.......} \quad (5.8)$$

Where: (n) is the number of training samples to be evaluated in the training phase, (Ai) is the actual output of the training sample, and (Pi) is the predicted output for the same training sample.

Moreover, the average percentage of difference (%Error) between the Predicted cost (Ai) and the Actual cost (Pi) was calculated for training set for monitoring process and also calculated for testing set for validating the model, according to equation (5.9). Figure (5.7) shows the calculation of error in the model.

$$\%\ Error = \frac{(Pi-Ai)}{Ai} x100 \quad \text{.......} \quad (5.9)$$

Step7 :Error calculations

Set	No.	Predicted cost	Actual Cost	Error	% Error	S.Error
TR	1	69,373,769.10	70,000,000.00	-626230.9048	0.9%	3.92165E+11
TR	2	67,794,173.38	67,000,000.00	794173.3768	1.2%	6.30711E+11
TR	3	5,848,418.77	6,000,000.00	-151581.2316	2.5%	22976869776
TR	4	12,654,532.32	15,000,000.00	-2345467.676	15.6%	5.50122E+12
TR	5	26,827,862.65	27,000,000.00	-172137.3452	0.6%	29631265601
TR	6	328,805,457.09	330,000,000.00	-1194542.914	0.4%	1.42693E+12
TR	7	29,214,531.03	25,000,000.00	4214531.028	16.9%	1.77623E+13
TR	8	35,203,014.23	35,000,000.00	203014.2278	0.6%	41214776680
TR	9	251,859.84-	6,000,000.00	-6251859.835	104.2%	3.90858E+13
TR	10	10,165,129.41	8,000,000.00	2165129.412	27.1%	4.68779E+12
TR	11	131,391,859.36	130,000,000.00	1391859.361	1.1%	1.93727E+12
TR	12	43,432,739.73	43,000,000.00	432739.7278	1.0%	1.87264E+11
TR	13	7,426,440.03	7,000,000.00	426440.026	6.1%	1.81851E+11
TR	14	15,218,197.55	16,000,000.00	-781802.4488	4.9%	6.11215E+11
TS	15	159,428,986.30	35,000,000.00	124428986.3	355.5%	1.54826E+16
TS	16	10,745,603.30	11,000,000.00	-254396.7037	2.3%	64717682839
TS	17	5,102,213.87	5,000,000.00	102213.8671	2.0%	10447674628
TS	18	80,590,843.16	80,000,000.00	590843.1621	0.7%	3.49096E+11

Figure (5.7): Error Calculation

8. Learning Role For Optimization Goal

The final step in the model execution is defining the learning role for the optimization purpose; this role was assigned, in the Evolver, to minimize the summation of the previously calculated RMSE for the training set. Moreover, additional two cells were assigned to calculate the average value for the percentage of error (% Error) for training set and testing set. Then, these values were used to monitor the change during the model learning. Figure (5.8) shows the cells assigned for learning role and monitoring.

RMSE	Training Error %	Testing Error %
6928981.708	13.1%	90.2%

Figure (5.8): The Assigned Cells for Learning Role and Monitoring

5.4 Model Implementation

After finishing of the Model Execution stage, it's quite important to focus on the settings adjusted in the Evolver before starting simulation. Firstly, all cells designated for input connecting weights were attributed in the Evolver as adjustable cells with decimal fraction values to allow Evolver to search for the best weight values, similarly all cells designated for output connecting weights were also attributed in the Evolver the same way as cells of input connection weight. Settings of the adjustable cells, such as mutation and crossover were kept in default value (0.1 & 0.5) respectively.

By the end of all previous steps and settings assignment, it is time to run the model and start optimization and gaining results.

5.5 Trial and Error Practices

To verify this research work, trial and error practices were carried out to conclude the best model. Thirty six trials were applied for model training. These trials were performed in two different groups as shown in table (5.2).The table represents a complete summary of description for both group of trials, It is divided into seven fields; Trials group number, description of the group, number of hidden layers, number of nodes in each layer, activation function for each hidden layer, exit activation function for output layer and the total number of trials in each group.

Table (5.2): Summary of Trials and Error Practices

Group Number	Description	Number of Nodes in Each Hidden Layer	Activation Function for Each Hidden Layer(s)	Exit Activation Function	Total Number of Trials
Group (1)	One hidden layer	• 8 nodes	• Tanh • Sigmoid • Linear	• Tanh • Sigmoid • Linear	9
Group (2)	Two hidden layers	• eight nodes for first one • 4 nodes for second one	• Tanh • Sigmoid • Linear	• Tanh • Sigmoid • Linear	27
Total					**36**

Table (5.3) shows the detailed log of all trials. It contains the number of hidden layers in each trial, the number of hidden nodes in each layer, activation function of each hidden layer, exit activation function in the output layer, root mean square error RMSE and the training Error (%Error).

Table (5.3): Detailed Log for Trials and Error Practices

Trial No.	No. of hidden Layers	No. of hidden Nodes	1st Layer Activation Function	2nd Layer Activation Function	Exit Function	RMSE	% Error
1	1	8	Tanh	N/A	Tanh	2510157.91	19.0%
2	1	8	Tanh	N/A	Sig	1570859.100	49%
3	1	8	Tanh	N/A	Linear	1151929.69	11.1%
4	1	8	Sig	N/A	Tanh	3415072.69	27.9%
5	1	8	Sig	N/A	Sig	1570859.100	49%
6	1	8	Sig	N/A	Linear	476374.08	13.1%
7	1	8	Linear	N/A	Tanh	4707821.18	41.1%
8	1	8	Linear	N/A	Sig	1570859.100	49%
9	1	8	Linear	N/A	Linear	1570859.100	49%
10	2	8 & 4	Tanh	Tanh	Tanh	2072249.63	16.1%
11	2	8 & 4	Tanh	Tanh	Sig	1570859.100	49%
12	2	8 & 4	Tanh	Tanh	Linear	527869.96	12.6%
13	2	8 & 4	Tanh	Sig	Tanh	2480541.99	23.0%
14	2	8 & 4	Tanh	Sig	Sig	1570859.100	49%
15	2	8 & 4	Tanh	Sig	Linear	3181757.65	27.9%
16	2	8 & 4	Tanh	Linear	Tanh	4521030.52	35.5%
17	2	8 & 4	Tanh	Linear	Sig	1570859.100	49%
18	2	8 & 4	Tanh	Linear	Linear	543916.28	12.2%
19	2	8 & 4	Sig	Tanh	Tanh	1381981.08	22.3%
20	2	8 & 4	Sig	Tanh	Sig	1570859.100	49%
21	2	8 & 4	Sig	Tanh	Linear	3577819.94	31.7%
22	2	8 & 4	Sig	Sig	Tanh	1181066.30	31.5%
23	2	8 & 4	Sig	Sig	Sig	1570859.100	49%
24	2	8 & 4	Sig	Sig	Linear	1260278.83	34.4%
25	2	8 & 4	Sig	Linear	Tanh	5043942.99	46.2%
26	2	8 & 4	Sig	Linear	Sig	1570859.100	49%
27	2	8 & 4	Sig	Linear	Linear	1011171.79	29.5%
28	2	8 & 4	Linear	Tanh	Tanh	4504308.19	37.7%
29	2	8 & 4	Linear	Tanh	Sig	1570859.100	49%
30	2	8 & 4	Linear	Tanh	Linear	1873278.30	41.2%
31	2	8 & 4	Linear	Sig	Tanh	1124561.11	22.9%

Trial No.	No. of hidden Layers	No. of hidden Nodes	1st Layer Activation Function	2nd Layer Activation Function	Exit Function	RMSE	% Error
32	2	8 & 4	Linear	Sig	Sig	1570859.100	49%
33	2	8 & 4	Linear	Sig	Linear	3873127.00	45.1%
34	2	8 & 4	Linear	Linear	Tanh	5833422.80	47.7%
35	2	8 & 4	Linear	Linear	Sig	1570859.100	49%
36	2	8 & 4	Linear	Linear	Linear	6581432.48	116.3%

As shown in table (5.3), the minimum RMS was concluded in trial number (6), this trial was clouded by gray in the table. Therefore, it is the recommended structure which should be tested. This structure consists of one hidden layers with activation function Sigmoid for summation of weighted Inputs, where number of hidden nodes were 8, whilst the exit function for output node was "Linear". In addition, the average absolute percentage of error (%Error) was (13.1%). Figure (5.9) shows the Graphically RMSE for each trial.

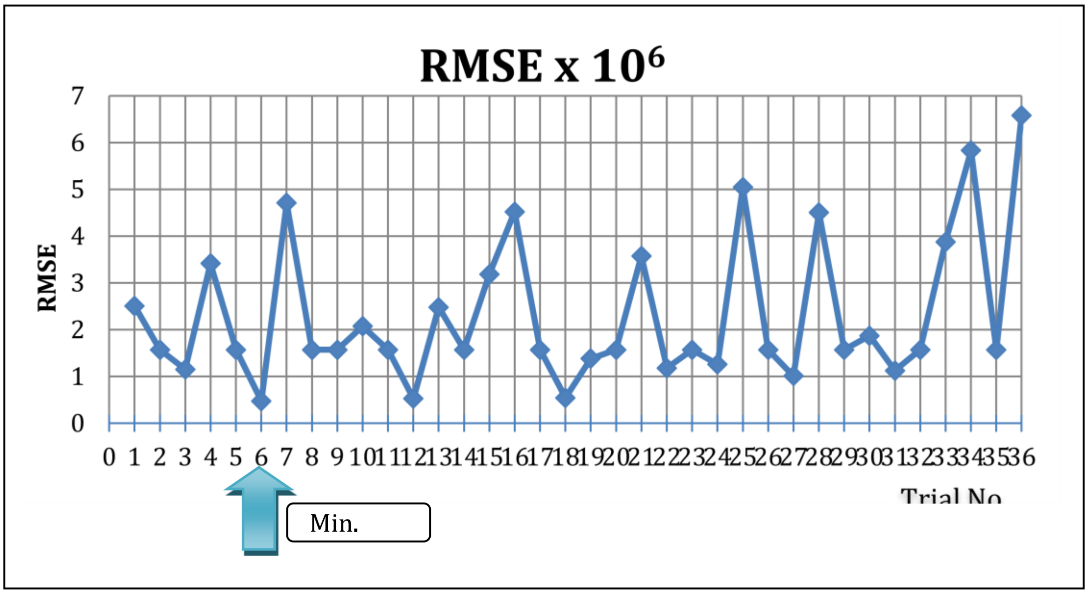


Figure (5.9): RMSE for Each Trial

5.6 Model Testing

Testing the network is essentially the same as training it, except that the network is shown facts it has never seen before, and no corrections or modifications are required. It is important to evaluate the performance of the network after the training process. If the results are good, the network will be ready to use. If not, this needs more or better data or to redesign the network. A part of the facts around 20%, i.e. four facts are set aside randomly from training facts. These facts are used to test the ability of network to predict a new output. Table (5.4) presents the actual cost and predicted cost for testing facts which are calculated using the developed model. It shows that the percentage of absolute difference of predicted cost (%Error) ranges from 0.7% to 2.3% with average value of 1.8% which is less than the previously mentioned average absolute percentage of error for the training facts (13.1%). Consequently, the model testing was successfully passed and it is accepted to be used in cost estimating processes for such type of projects that are containing sterile buildings.

Table (5.4) : Testing Results of Developed Neural Networks Model

Project No.	Actual Cost	Predicted Cost	Difference	Absolute Difference	% Error
1	35000000	35776871.77	776871.7691	776871.7691	2.2%
2	11000000	10745603.3	-254396.7037	254396.7037	2.3%
3	5000000	5102213.867	102213.8671	102213.8671	2.0%
4	80000000	80590843.16	590843.1621	590843.1621	0.7%

5.7 Model Validation

Now it is the time to validate the developed model to prove that the model is reliable to be used in future in predicting process for the new projects in the early stage of its life cycle. This validation will be done based on a new two projects constructed earlier, the archived data of these two projects were presented in appendix (B).

The validation process was done using the same criteria of trial no. (3); structure, activation function and connection weights. Table (5.5) shows the result obtained from validation, where the % Error was (8%) for the first project and (6.28%) for the second on. These results prove certainly that the developed model is accurate and reliable enough to be used as a predicting tool at the conceptual stage, moreover, at all project stages.

Table (5.5): Validation Results of Developed Neural Networks Model

Project No.	Actual Cost	Predicted Cost	Difference	Absolute Difference	% Error
1	97,000,000	104,789,233	-7,789,233	7,789,233	8%
2	150,000,000	159,428,986	-9,428,986	9,428,986	6.28%

5.8 Model Limitation

The developed model is limited to prediction of final construction cost of the industrial construction projects, such as pharmaceutical, food and dairy projects, it can't be used for any other type of projects without checking the validity of this model to predict a reliable results.

5.9 Conclusion

This chapter presents the methodology used to build the parametric cost estimation model for pharmaceutical and food projects. Moreover, the developed model was tested and validated to prove that the model is capable to predict the construction cost for future projects with acceptable accuracy. Accordingly, the resultant percentage of error (% Error) can be added to as a cost contingency to the predicted project cost.

CHAPTER 6

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

CHAPTER 6

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

6.1 Introduction

This thesis is presented in six Chapters encircling the whole research essence. This Chapter presents the summary of the study, ultimately divulging the digest of major findings drawn from the study. The research findings were presented while considering the final model, and their interpretations are also briefly mentioned. This Chapter also cites the appropriate recommendations, which the researcher has developed based on the conclusions of the thesis. Some recommendations for further studies in the same area were also indicated and suggested according to the perception of the researcher.

6.2 Summary

Cost estimating is one of the key inputs for project's success. In the early stage of project's life cycle, providing a reliable cost estimation isn't an easy task. In such stage, there is a lack of accurate and/or complete project data. This effort focus is directed to parametric construction cost estimating for pharmaceutical and food projects that enclose such type of sterile buildings.

Project's parametric cost is greatly affected by many uncertain but predictable factors. In this thesis, the important cost factors affecting construction cost were identified based on a comprehensive questionnaire survey distributed among a sample of expert engineers in Egypt. Moreover, the questionnaire results have been extensively analyzed to determine the most important cost factors. Moreover, a neural networks model was developed in order to help project's cost estimators to have a more reliable and easy cost estimating.

6.3 Conclusion

Throughout the entire effort conducted for this effort, the following conclusions were drawn:

- Based on literature review and interviews done with experts, it was found that thirty seven factors may affect the expected construction cost of such type of projects; fourteen factors were originated from the literature review and the remaining twenty four from the interviews.

- Based on a questionnaire survey among the different construction experts in Egypt, the most important cost indicators for industrial construction projects which comprise sterile were identified as following:

  1. Currency exchange rate during the month of studying the project.
  2. Egypt Consumer price index during the month of studying the project.
  3. Accumulative built-up area (total area).
  4. Accumulative Sterile areas (total area).
  5. Other supplementary buildings (water tank, administration, warehouse, … etc) (total area).
  6. Project location.
  7. Target market (regional or international).
  8. International insurances (if any).
  9. Desired level of contractor's prequalification.
  10. Desired completion time for the project.
  11. Desired structural system.
  12. Buildings closeness (attached, semi-attached or separated).
  13. Project Status (renovation, extension or….etc).

- Pertinent data regarding eighteen real life building construction projects were collected from the Egyptian construction field. Eventually, such projects were divided into two groups. First, fourteen of these projects were considered in developing the proposed neural networks model. Second, four projects were exploited for testing the validity of that model.

- The Artificial Neural Networks model was introduced as a management tool that can enhance current automation efforts in the construction industry. ANN model was prepared in order to predict the expected construction cost of any future project in the early stage of the project life cycle, especially in the conceptual phase. Thirty six trials were applied to identify the best structure of the proposed model. The validity of the proposed model was tested through the last four projects facts. The results of this test clearly shows that the percentage of absolute difference of predicted cost (%Error) ranges from 0.7% to 2.3% with average value of 1.8% which is less than the previously mentioned average absolute percentage of error for the training set (13.1%). Therefore the model testing has successfully passed and it is valid to be used in cost estimating process.

6.4 Recommendations for Future Studies

This subject opens the door for a lot of future researches. The following potential areas of studies, if explored, would provide increased validity to the findings of this thesis:

- For cost indicator number nine (I9) for case of renovation, it is suggested to study separately the effect of building age and the type of production on the required cost for renovation due to the consequence of ageing and building deterioration respectively, it is also recommended to split this factor as a new indicator.

- The model should be augmented to take into consideration the other different types of construction projects. For example: the medical, commercial and administrative construction projects.

- The development of artificial neural network models or any other technique requires the presence of structured and well organized database for the completed projects in the construction firms. Unfortunately, most of the Egyptian construction firms have no structured database system that can provide researchers with the required information. It is recommended that a standard database system for storing information about all completed projects should be developed and applied by the construction companies in Egypt.

REFERENCES

REFERENCES

1) Adeli, H., & Wu, M. Y. (1998). "Regularization Neural Network for Construction Cost Estimation." Journal of Construction Engineering and Management-ASCE, 124(1): PP 18-24.

2) Al-Tabtabai, H., Kartam, N., Flood, I, and Alex, A.P. (2008). "Expert Judgment in Forecasting Construction Project Completion." Engineering Construction and Architectural Management, 4(4): PP 271-293.

3) Arafa, M. and Alqedra, M. (2011). "Early Stage Cost Estimation of Buildings Construction Projects Using Artificial Neural Networks." Journal of Artificial Inelegance, 4(1): PP 63 - 75.

4) Arafa, M. and Alqedra M. (2011). "Early Stage Cost Estimation of Buildings Construction Projects Using Artificial Neural Networks." Journal of Artificial Intelligence 4(1): PP 63-75.

5) Asal, E. (2014). "Factors Affecting Building Construction projects' Cost Estimating." M.Sc. Thesis. College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Cairo, Egypt.

6) Attal, A. (2010). "Development of Neural Network Models for Prediction of Highway Construction Cost and Project Duration." M.Sc. Thesis. College of Engineering and Technology, Ohio University.

7) Ayed, A. (1997). "Parametric Cost Estimating of Highway Projects Using Neural Networks." M.Sc. Thesis. Faculty of Engineering & Applied Sciences, Memorial University, St. John's, Newfoundland, Canada.

8) Bouabaz, M., and Hamami, M. (2008). "A Cost Estimation Model for Repair Bridges Based on Artificial Neural Network." American Journal of Applied Sciences, 5(4): PP 334 – 339.

9) Charles, A. T. (2006). "A Bottoms-Up Approach To Cost Estimation Using Parametric Inputs." M.Sc. Thesis. College Of Engineering and Technology, Ohio University, United States of America.

10) Creese, R., and Li, L. (2005). "Cost Estimation of Timber Bridges Using Neural Networks. Cost Engineering." AACE International, 37(2): PP 14-18.

11) El-Gafy, M. A. (2001). "Neural Network Model for Parametric Cost Estimation of Sewer Projects." M.Sc. Thesis. Faculty of Engineering, Cairo University, Giza, Egypt.

12) Emsley, M., Lowe, D., Duff, A., Harding, A. and Hickson, A. (2002). "Data Modeling and The Application of A Neural Network Approach To The Prediction of Total Construction Costs." Construction Management and Economic, 20(6): PP 465- 472.

13) Emsley, M., Lowe, D., Duff, A., Harding, A. and Hickson, A. (2007). "Relationships Between Total Construction Cost and Design Related Variables." Journal of Financial Management of Property and Construction, 12(1): PP 11- 24.

14) Garza, J.,and Rouhana, K. (1995). "Neural Network Versus Parameter Based Application." Cost Engineering, AACE International, 37(2): PP 14-18

15) Bartlett, James (2001). "Determining Appropriate Sample Size in Survey Research." Information Technology, Learning, and Performance Journal, Vol. (19), No. (1): PP 43 – 50.

16) Gwang-Hee, K., Sang-Hoon, A., and Kyung, I. (2004). "Comparison of Construction Cost Estimating Models Based on Regression Analysis, Neural Networks, and Case-Based Reasoning." Journal of Building and Environment, 39: PP 1235 – 1242.

17) Hegazy, T., Moselhi, O., and Fazio, P. (1994). "Analogy-Based Solution To Markup Estimation Problem." J. Comp. in Civil Engineering, ASCE, 8(1): PP 72 - 87.

18) Hegazy, T.,Ayed,A. (1998). "Neural Network Model for Parametric CostEstimation of Highway Projects." Journal of Construction Engineering and Management-ASCE, 124(3): PP210-218.

19) Hosny, H. (2011). "Contingency Assessment in Time Schedule Using Artificial Neural Network." M.Sc. Thesis. College Of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Cairo, Egypt.

20) Jamshid, S. (2005). "Cost Estimation of Highway Projects in Developing Countries: Artificial Neural Network Approach." Journal of the Eastern Asia Society for Transportation Studies, 6: PP. 1036 – 1047.

21) Lekan, A. M. (2011). "Neural Network-Based Cost Predictive Model for Building Works." Ph.D. Thesis. School of Environmental Sciences, College Of Science and Technology, Covenant University, Ota, Ogun State.

22) Mckim, R. (2005). "Neural Networks and Identification of Estimating Risks." AACE Transactions, 35 (8): PP 52-56.

23) Meijer, B. (2009). "Neural Networks for Device and Circuit Modeling." I.E.E.E Transaction on Circuits and Systems, 37: PP 335-346.

24) Morcous, G. (1997). "Estimating of Quantities and Cost of P.C. Bridges Over The Nile in Egypt: Conventional and Neural Networks Approaches." M.Sc. Thesis. Faculty of Engineering, Cairo University, Giza, Egypt.

25) Murat, H., and Dogan, S. (2004). "A Neural Network Approach for Early Cost Estimation of Structural Systems of Buildings." International Journal of Project Management, 22(7): PP 595 - 602.

26) Phaobunjong, K.. (2002). "Neural Parametric Cost Estimating Model for Conceptual Cost Estimating of Building Construction Projects." Ph.D. Thesis. Faculty of the Graduate School, Texas University, Austin.

27) Project Management Body of Knowledge (PMBOK® Guide). (2008). Fourth Edition. PP 165-188

28) Salam, A., Defersha, F., Bhuiyan, N., and Chen, M. (2012). "A Case Study on Target Cost Estimation Using Back-Propagation and Genetic Algorithm Trained Neural Networks." Journal of Cost Analysis and Parametrics, 5(2): PP 87-97.

29) Setyawati, B. R., Sahirman, S. and Creese, R. C. (2002). "Neural Networks for Cost Estimation." ASCE International Transactions, EST.13(9): PP 22 - 30.

30) Siqueira, I. (1999). "Neural Network-Based Cost Estimating." M.Sc. Thesis. Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, Quebec, Canada.

31) Stoy, C., Pollalis, S., and Schalcher, H. (2008). "Drivers for Cost Estimating in Early Design: Case Study of Residential Construction." Journal of Construction Engineering and Management, 134(1): PP 32- 39.

32) Tunc, M. (2003). "Computerized Cost Estimation for Forging Industry." M.Sc. Thesis. School of Natural and Applied Sciences, Department Of Mechanical Engineering, Middle East Technical University, Ankara, Turkey.

33) Wheaton, C., and Simonton, E. (2007). "The Secular and Cyclic Behavior of 'True' Construction Costs." Journal of Real Estate Research, 20(1): PP 1- 25.

APPENDIXES

APPENDIXES

APPENDIX (A)

I. Sample of Filled Questionnaire

ARAB ACADEMY FOR SCIENCE, TECHNOLOGY

& MARITIME TRANSPORT

COLLEGE OF ENGINEERING & TECHNOLOGY

DEPARTMENT OF CONSTRUCTION & BUILDING

ENGINEERING

MASTER DEGREE RESEARCH

Title:

Parametric Cost Estimating of Sterile Building Using Artificial Neural Network & Genetic Algorithm Model

Questionnaire

Introduction

This Questionnaire is a part of Data collection Chapter in this research. The results from the completed questionnaire should be used as guidelines in prediction of construction cost for future projects duration. Any participant is kindly requested to complete this questionnaire in respect of his/her experience in construction of industrial Projects in Egypt. It should be noted that civil, bridges or industrial projects are not included in this research.

Thanks for your Participation and Cooperation.

Researcher Data

Name: Mohamed SaadAllah Zahran

Job title: Project Manager

Organization: Engineering Consultants Group ECG

Phone: +2 (0100) 242 0017

E-mail: **Mohamed.zahran@ecgsa.com**

Response to the Questionnaire

In this questionnaire, the respondent should determine the impact and the degree of existence for each factor based on his/her experience in the field of industrial construction projects.

Impact of any cost factor means how strongly the cost affected by the change of that factor, while the degree of existence of the factors represent the degree of repetitiveness of this factor.

The response should be in regards of five degree scale ranges from very low to very high.

Cost Estimation – Questionnaire											
Definition Cost Factors are the contributory factors affecting the total cost of projects											
No.	**Cost Factors**	**Impact**					**Degree of Existence**				
		V.L	L	M	H	V.H	V.L	L	M	H	V.H
A	**General Factors** General Factors represent the contributory factors related to the initial Project Data, Constraints and owner requirements										
1	Project location				√				√		
2	Site accessibility		√					√			
3	Site Constraints		√					√			
4	Desired completion time for the project					√				√	
5	Subsistence of time constrains					√					√
6	Owner requirements for bid packaging for multiple contractors			√					√		
7	Environmental impact assurance system requirements			√						√	
8	Applying safety system during construction		√							√	
B	**Engineering related Factors** Engineering related factors represents the technical factors that can be calculated, assumed or demonstrated by numbers										
1	Site topography			√					√		
2	Accumulative built-up area					√					√
3	Buildings closeness (attached, semi-attached or separated)			√					√		
4	Other supplementary buildings (W.tank, administration, warehouse,… etc)				√					√	
5	Accumulative Sterile Areas (total area)					√					√
6	Structural design loads				√				√		
7	Geotechnical nature of soil			√					√		
8	Desired structural system				√						√
9	Desired HVAC system					√					√
10	Desired Firefighting system			√							√
C	**Resources related Factors** Resources factors are the factors that are directly related to the required amount of resources, their skill level and/or price										
1	Consultant fees		√							√	
2	Desired level of contractor's prequalification				√					√	

3	Contractor overhead				√					√	
4	Need for special contractor(s)				√					√	
5	Reinforcement price					√					√
6	Cement price			√					√		
7	stainless Steel price					√					√
8	Labor price				√					√	
9	% of imported material					√					√
10	Availability of required power			√					√		
D	**Special factors** Special Factors are the mandatory factors related to the local and/or international regulations and requirements for factory production										
1	Target market (regional or international)					√				√	
2	Type of products and type of production method				√					√	
3	Special finishing required for sterile areas					√					√
4	Additional requirements for structural system regarding sterile manufacturing				√						√
5	Additional requirements for HVAC system regarding sterile manufacturing				√						√
6	Industrial safety requirements (firefighting, fire alarm, .. etc)				√						√
E	**External Factors** External factors represents the factors related to Local and international economy status										
1	Currency exchange rate				√						√
2	Inflation				√					√	
3	International insurances (if any)				√					√	

Recipient Data

Company name:	Howeedy Consultant
Recipient name:	Hisham Howeedy
Recipient Title:	General Manager
Recipient Experience:	5 ☐ 10 ☐ 15 ☐ 20 ☐ 25 ■
Signature:	
Date:	25-11-2013

Key

	V.L	**L**	**M**	**H**	**V.H**
	Very Low	Low	Moderate	High	Very High

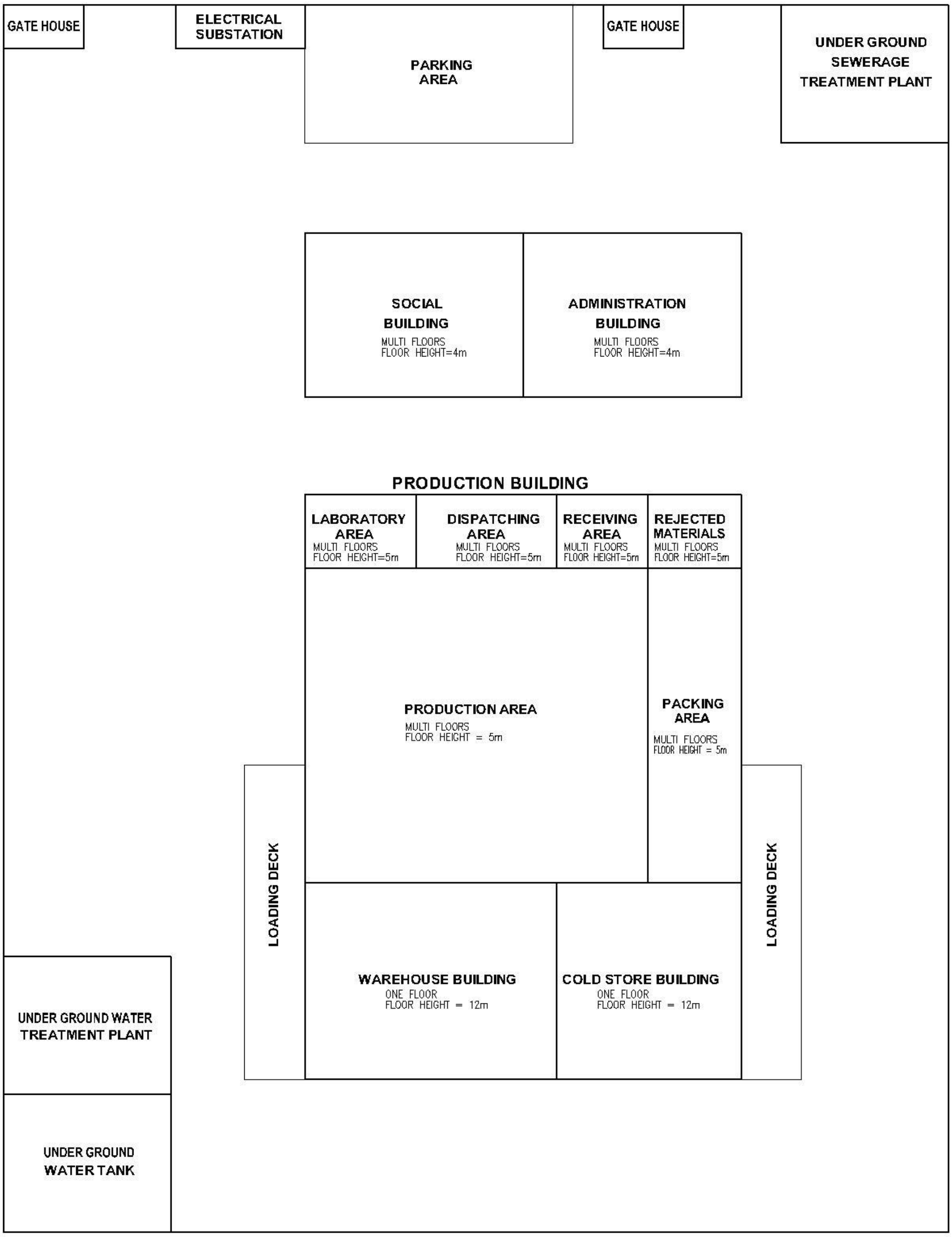
II. List of Respondents

No	Respondent Name	Respondent Title / Position	Name of Org.	Years of Exp.
1	Mohamed Samy El Komy	Project Manger	ECG	15
2	Nirven Yaseen	Deputy Project Manager	ECG	10
3	Islam Shawkey	Project Manager	ECG	10
4	Mohamed El Berelossy	Project Manager	ECG	10
5	Ahmed Said	Deputy Project Manager	ECG	10
6	Hisham Koraitim	Project Manager	ECG	20
7	Ahmed Abd El Bar	Project Manager	ECG	25
8	Ahmed Abdel Moanem Yehia	Cost Manager	ECG	7
9	Aya Diaa	Cost Engineer	ECG	8
10	Yamen Omar	Deputy Project Manager	ECG	10
11	Nayera Anis	Project Engineer	ECG	25
12	Amr Abu El Neel	Project Manager	ECG	10
13	Mohamed Mehrez	Project Manager	ECG	15
14	Mohamed Safwat	Cost Engineer	ECG	9
15	Mohamed Mohamed Islam	Cost Engineer	ECG	8
16	Ramy El Khalafawy	Project Manager	ECG	10
17	Ahmed Samy Hassan	Project Manager	ECG	10
18	Mohamed Ezzat Mehrez	Deputy Project Manager	ECG	15
19	Mahmoud Elsheemy	Sr. Cost Engineer	ECG	15
20	Ahmed Nayel	Sr. Cost Engineer	ECG	10
21	Mahmoud Hassaan	Cost Engineer	ECG	8
22	Dalia Hussain	Sr. Cost Engineer	ECG	10
23	Mohamed Saleh	Project Manager	ECG	20
24	Mohamed Sayed	Senior Structural Engineer	HC	10
25	Mostafa Helmi	Design Office Manger	HC	25
26	Emam Ibraheim	Head Of Architectural Department	HC	25
27	Hatem Fayez	Project Manger	HC	20
28	Tamer Selim	Senior Structural Eng. - Technical Office	HC	20
29	Hisham Amin Howeedy	Member Of Board	HC	25
30	Ahmed Abd El Azim Shokry	Project Manager	HC	25
31	Hussein Mostsfa	Head Of Electrical Department	HC	25

No	Respondent Name	Respondent Title / Position	Name of Org.	Years of Exp.
32	Mostafa Abd El Magid	Project Manager	HC	25
33	Emad Abdel Aziz	Project Manager	HC	20
34	Kamal Moustafa	Project Manager	HC	20
35	Mahmoud Maged	Project Manager	HC	20
36	Hesham Thabet	Project Manager	HC	20
37	Mohamed Anwar	Project Manager	VBS	20
38	Taher Abdel Aziz	CEO	VBS	25
39	Amira Salah Eldin	Sr. Project Management Consultant	VBS	15
40	Tarek El Menayar	Sr. Civil Engineer	VBS	10
41	Ahmed Zahran	Project Manager	ZDC	20
42	Ehab Darwish	Project Manager	ZDC	20
43	Sayed Lashin	Project Manager	ZDC	20
44	Hassan Thabet	Project Manager	Juhaina	25
45	Ahmed El Nimr	Head of Engineering Department	Novartis	25
46	Hassan El Hussainy	Project Manager	Novartis	20
47	Hossam Gamal	Project Manager	Novartis	20
48	Freddy Marcos	Projects Manager	Nestle	25
49	Moustafa Hafez	Project Manager	Nestle	20
50	Rob Johns	Project Manager	Nestle	20
51	Khaled Rakha	Head of Cost Estimation Department	Hill Intl	25
52	Tamer Haggag	Project Manager	Hill Intl	20

APPENDIX (B)

I. Schematic Description for Targeted Projects



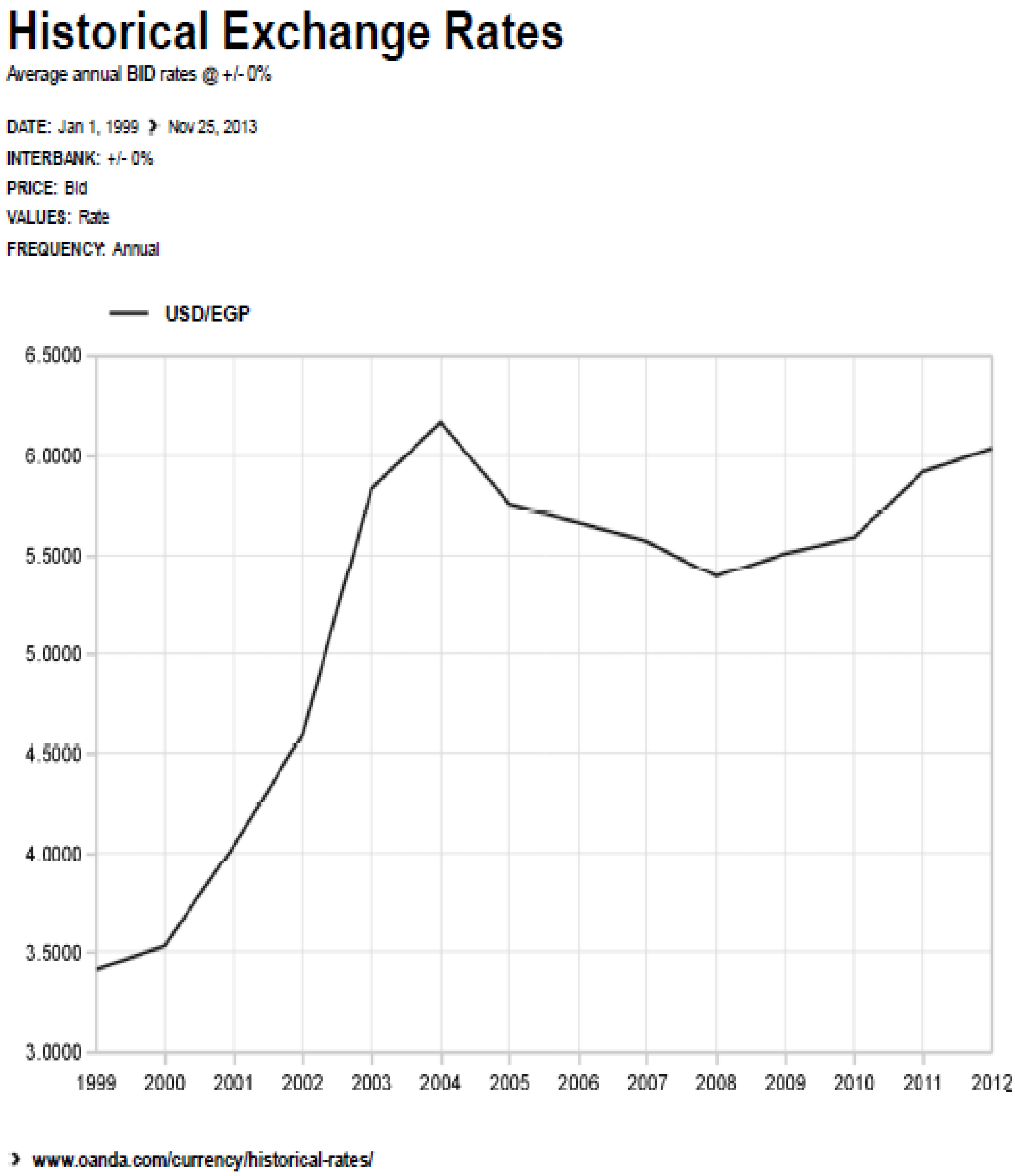
Each project consists of number of buildings as presented in the above schematic layout. Generally in such projects, there are common characteristics between the comparable buildings. These characteristics are in the following table.

Building Name	Number of Floors	Floor Height (m)	Hygienic Case
Administration Building	Multi Story	4	No
Social Building	Multi Story	4	No
Cold store Building	Single Story	12	Hygienic
Warehouse Building	Single Story	12	Hygienic
Production Building	Multi Story	5	Sterile
Gate House	Single Story	3	No
Electrical Substation	Single Story	3	No
Water Tank	Single Story	4.5	No
Water Treatment Plant	Single Story	4	Sterile
Sewerage Treatment Plant	Single Story	4	No

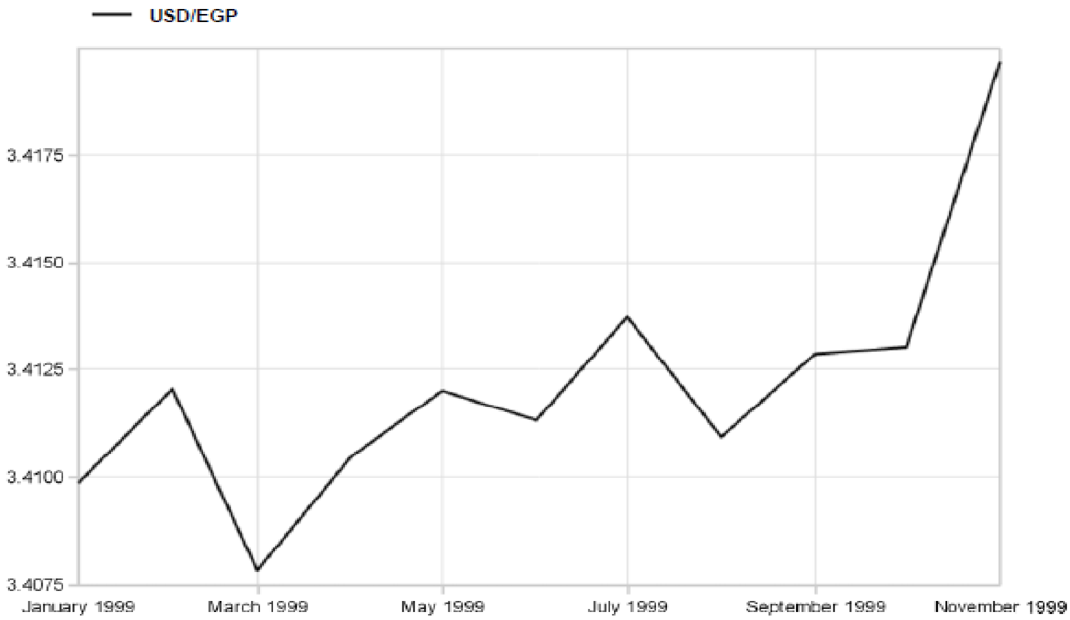
II. The list of historical data collected from two new projects constructed in Egypt.

No.	**1**	**2**
Project (I_X)	**Nerhadou Factory**	**Doehler Factory**
I_1 (LE)	6.4	6.4
I_2	127.9	127.9
I_3 (Month)	11	12
I_4 (m^2)	6000	19000
I_5 (m^2)	1200	1500
I_6 (m^2)	4850	2500
I_7	Mixed	Mixed
I_8	Smi-attached	Attached
I_9	Entirely new project	new building in existing project
I_{10}	Giza	Giza
I_{11}	Regional market	Regional market
I_{12}	No	No
I_{13}	Moderate	High
Cost (LE)	97,000,000	150,000,000

III. The Collected Historical Data of the Exchange Rate



DATE: Jan 1, 1999 > Dec 30, 1999
INTERBANK: +/- 0%
PRICE: Bid
VALUES: Rate
FREQUENCY: Monthly

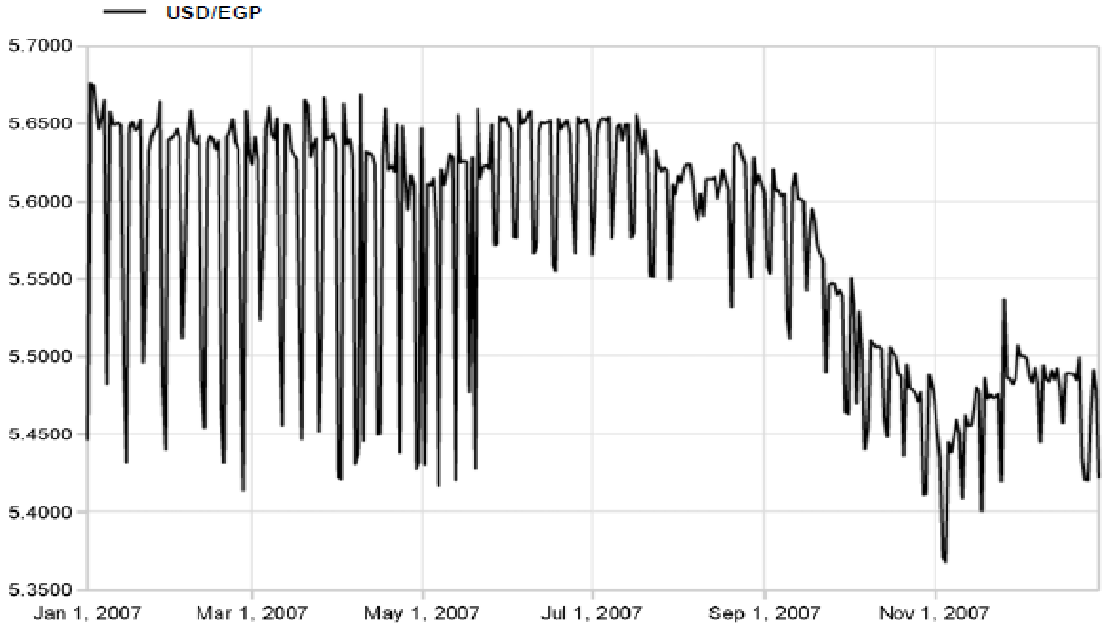


> **www.oanda.com/currency/historical-rates/**

DATE: Jan 1, 2003 > Dec 30, 2003
INTERBANK: +/- 0%
PRICE: Bid
VALUES: Rate
FREQUENCY: Daily

> **www.oanda.com/currency/historical-rates/**

DATE: Jan 1, 2007 › Dec 30, 2007
INTERBANK: +/- 0%
PRICE: Bid
VALUES: Rate
FREQUENCY: Daily



› www.oanda.com/currency/historical-rates/

DATE: Jan 1, 2009 › Dec 31, 2009
INTERBANK: +/- 0%
PRICE: Bid
VALUES: Rate
FREQUENCY: Weekly

› www.oanda.com/currency/historical-rates/

DATE: Jan 1, 2011 › Dec 31, 2011
INTERBANK: +/- 0%
PRICE: Bid
VALUES: Rate
FREQUENCY: Daily

› **www.oanda.com/currency/historical-rates/**

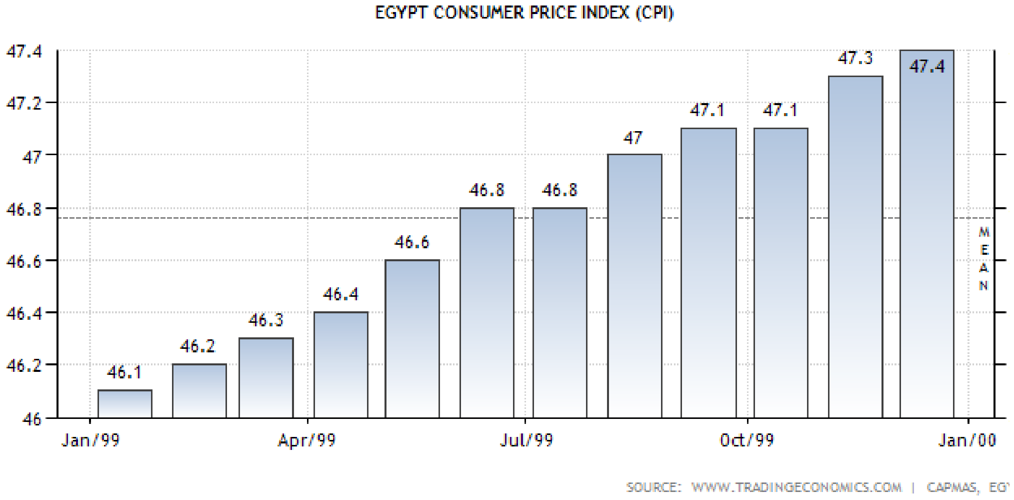
DATE: Jan 1, 2012 › Dec 31, 2012
INTERBANK: +/- 0%
PRICE: Bid
VALUES: Rate
FREQUENCY: Weekly

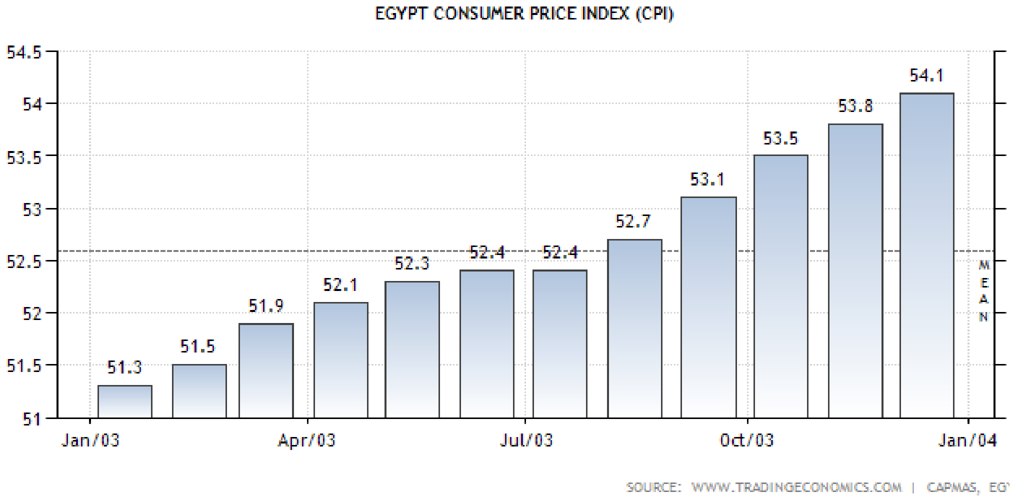
› **www.oanda.com/currency/historical-rates/**

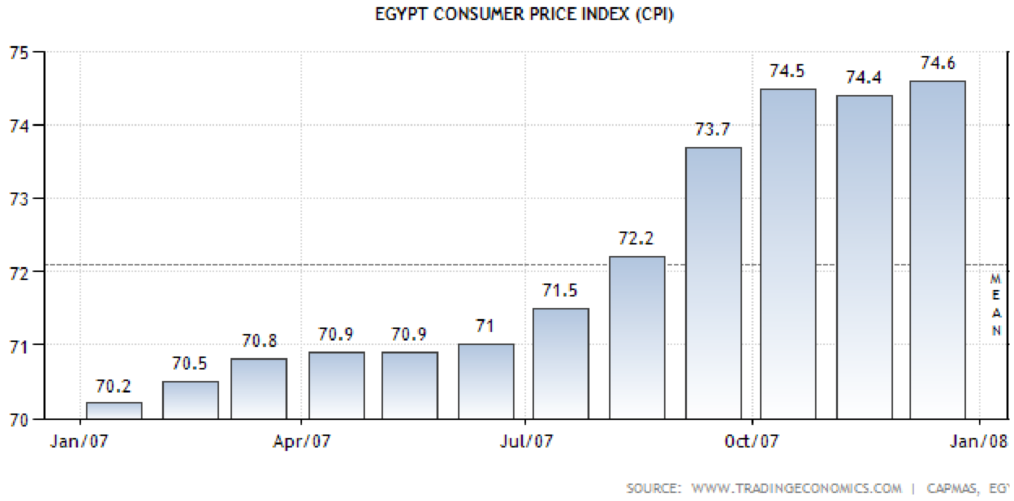
DATE: Jan 1, 2013 ➤ Nov 25, 2013
INTERBANK: +/- 0%
PRICE: Bid
VALUES: Rate
FREQUENCY: Weekly

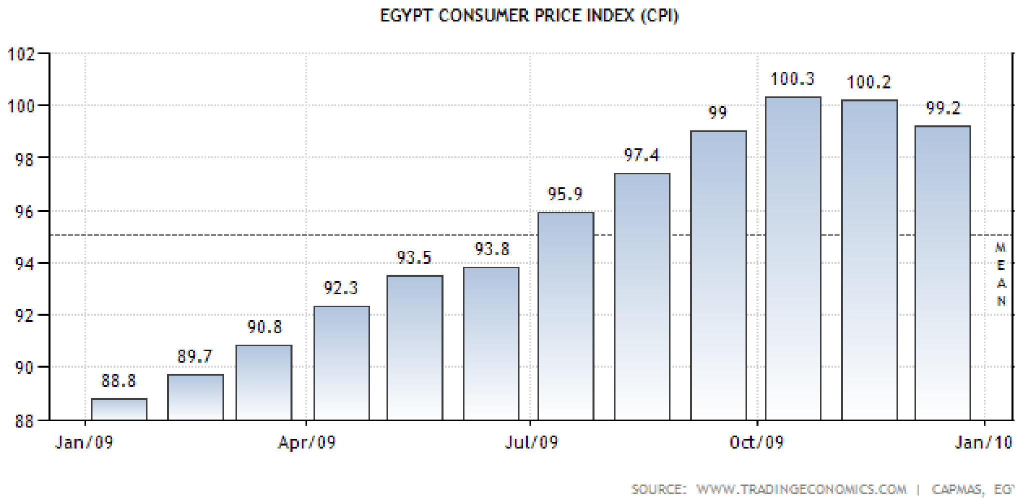
➤ www.oanda.com/currency/historical-rates/

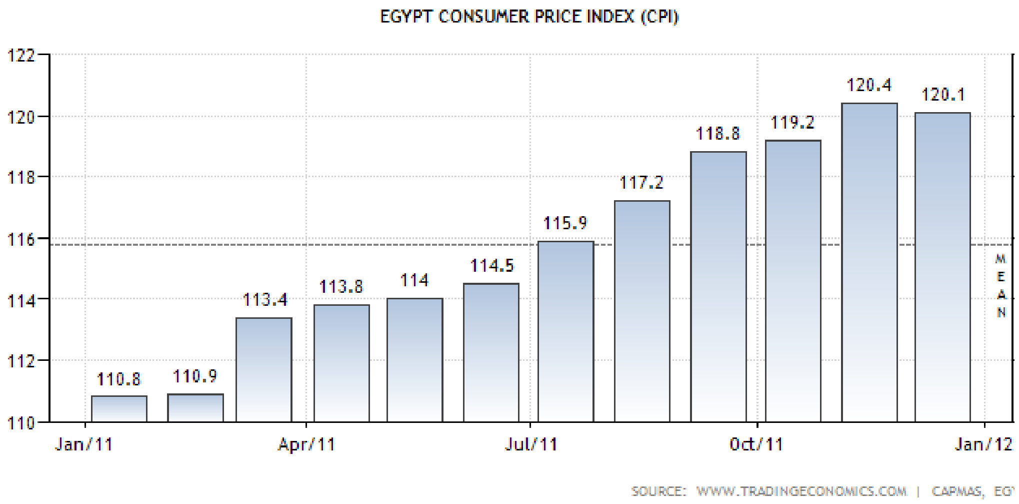
IV. The Collected Historical Data of Consumer Price Index

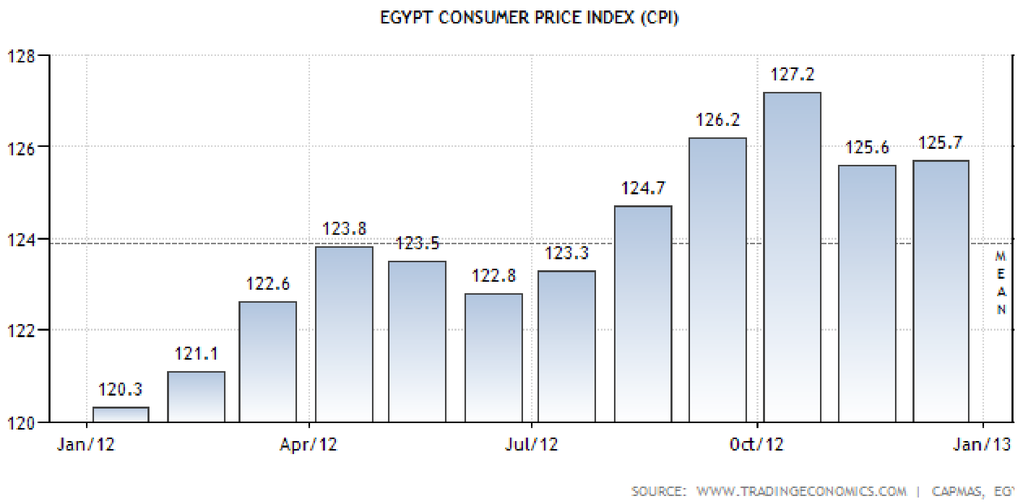







EGYPT CONSUMER PRICE INDEX (CPI)
102
100
98
96
94
92
90
88
88.8
89.7
90.8
92.3
93.5
93.8
95.9
97.4
99
100.3
100.2
99.2
MEAN
Jan/09
Apr/09
Jul/09
Oct/09
Jan/10
SOURCE: WWW.TRADINGECONOMICS.COM | CAPMAS, EG


EGYPT CONSUMER PRICE INDEX (CPI)
122
120
118
116
114
112
110
110.8
110.9
113.4
113.8
114
114.5
115.9
117.2
118.8
119.2
120.4
120.1
MEAN
Jan/11
Apr/11
Jul/11
Oct/11
Jan/12
SOURCE: WWW.TRADINGECONOMICS.COM | CAPMAS, EG


EGYPT CONSUMER PRICE INDEX (CPI)
128
126
124
122
120
120.3
121.1
122.6
123.8
123.5
122.8
123.3
124.7
126.2
127.2
125.6
125.7
MEAN
Jan/12
Apr/12
Jul/12
Oct/12
Jan/13
SOURCE: WWW.TRADINGECONOMICS.COM | CAPMAS, EG


EGYPT CONSUMER PRICE INDEX (CPI)
142
140
138
136
134
132
130
128
126
127.9
131.1
131.9
133.8
133.6
134.8
136
136.9
139
140.5
MEAN
Jan/13
Feb/13
Mar/13
Apr/13
May/13
Jun/13
Jul/13
Aug/13
Sep/13
Oct/13
Nov/13
SOURCE: WWW.TRADINGECONOMICS.COM | CAPMAS, EG

ملخص البحث

تعتبر التكلفة الأولية للمشروع من أكثر الامور الهامة خلال مرحلة تخطيط المشروع، حيث تتم عملية تقدير هذه التكلفة فى المراحل المبكرة من حياة المشروع حيث لا تتوافر معلومات وافية عن المشروع والتى بدورها تؤثر على تقدير تكلفة إنشاء المشروع. ويمكن تحديد أهمية التكلفة التقديرية للمشروع فيما بوجه عام كالتالي: (1) تاكيد مدى مصداقية دراسات الجدوى للمشروع، (2) تساعد أصحاب المصلحة من المشروع في اتخاذ القرارات المناسبة في الوقت المناسب، (3) تعطي تصور جيد عن التدفق النقدي للمشروع، (4) تعتبر حجر الاساس لعملية المراقبة والتحكم في تكلفة المشروع، (5) تقلل من المخاطر المحتملة في تكلفة المشروع. وبينما تم اثبات أهمية التكلفة التقديرية الاولية للمشروع بواسطة العديد من الابحاث، إلا أنه لا يوجد دراسات كافية لتحسين جودة هذه العملية ليسيرها و زيادة دقتها وذلك لإنشاء المشروعات الصناعية على وجه الخصوص.

الهدف الرئيسي من هذه الدراسة هو بناء نموذج دقيق يمكن الاعتماد عليه فى عملية تقدير التكلفة الاولية للمشروع، بحيث يمكن إستخدام هذا النموذج بكل سهولة ويسر من قبل الشركات أو الأفراد. هذا النموذج سوف يكون مخصص لنوع محدد من المشروعات الصناعية والتي تحتوى على مناطق معقمة مثل مصانع الأدوية والأغذية ومنتجات الالبان فقط ويقتصر استخدامه عليها.

من أجل بناء هذا النموذج تم تحديد العوامل الاكثر أهمية التي تؤثر على تكلفة المشروع عن طريق استبيان تم توزيعه على المتخصصين فى هذا النوع من المشاريع، ولقد استنتجت هذه دراسة الاستبيان أن العوامل الاكثر أهمية هى كالتالى: (1) سعر صرف العملة، (2) مؤشر أسعار المستهلك، (3) المساحة البنائية الاجمالية للمبنى، (4) المساحة الاجمالية للمناطق المعقمة، (5) المساحة الاجمالية للمبانى الملحقة، (6) موقع المشروع، (7) نوع السوق المستهدف، (8) وجود وثائق تامينية دولية، (9) المستوى الفنى للمقاول والذى يتم تحديده بواسطة المالك، (10) المدة المقررة للتنفيذ المشروع والتى يتم تحديدها أيضا عن طريق المالك، (11) النظام الانشائى المستخدم، (12) مدى تقارب أو التصاق المبانى مع بعضها البعض، (13) حالة المشروع؛ هل هو مشروع جديد أم إمتداد أو تجديد لمشروع قائم .

علاوة على ذلك، تم تصميم هيكل نموذج العمل بإستخدام هيكلة الخلايا العصبية الإصطناعيه، حيث تكون العوامل الاكثر أهمية للتكلفة هى المدخلات لهذا النموذج ويكون المخرج من هذا النموذج هو التكلقة التقديرية المتوقعة للمشروع. علاوة على ذلك، يمكن إستخدام متوسط قيمة الخطأ فى حساب قيمة احتياطية للتكلفة يمكن إضافتها علي القيمة التقديرية الناتجة من إستخدام النوذج لتجنب أي مخاطر مستقبلية.

تم بناء هذا النموذج كما تم تدريبه بإستخدام البيانات والمعلومات المجمعة من مشروعات تم تنفيذها سابقا وتتشابه مع بعضها البعض من حيث طبيعة ونوع المشروع، وقد كان عدد المشروعات المجمعة هو ثمانية عشر مشروعا ، تم تدريب النموذج علي أربعة عشر منها، بينما تم اختبار دقة النوذج علي أربعة منها وذلك لتحديد جودة النموذج، ومن ثم تم اختبار النموذج على مشروعين تم الإنتهاء من تنفيذهما لاحقا وذلك لتقييم دقة ومصداقية استخدام النموذج، ولقد أثبت عملية التقييم تلك دقة ومصداقية النموذج فى حالة استخدامه لتقدير التكلفة الاجمالية لتنفيذ المشروعات المستقبلية وذلك بدرجة من الدقة (8٪) وهى جديرة بالتقدير.

الأكاديمية العربية للعلوم والتكنولوجيا والنقل البحري

كلية الهندسة والتكنولوجيا

تقديرالتكلفة البارامتريه للمشروعات ذات المباني المعقمة باستخدام الشبكات العصبية الصناعية ونظرية التطور الجيني

رسالة ماجستير كاستيفاء جزئي لمتطلبات الحصول على درجة الماجستير في

هندسة التشييد والبناء

مقدمة من
م/ محمد سعدالله زهران
بكالوريوس الهندسة المدنية
كلية الهندسة – جامعة عين شمس

تحت إشرف

أ.د. عبدالمنعم ياسين سند
أستاذ قسم هندسة التشييد
كلية الهندسة والتكنولوجيا – الأكاديمية العربية للعلوم والتكنولوجيا والنقل البحري
القاهرة – مصر

أ.د. حسام الدين حسني محمد
أستاذ قسم هندسة التشييد
كلية الهندسة - جامعة الزقازيق
الزقازيق – مصر

أبريل 2015