The Dissertation Committee for Kan Phaobunjong Certifies that this is the approved version of the following dissertation:

PARAMETRIC COST ESTIMATING MODEL FOR CONCEPTUAL COST ESTIMATING OF BUILDING CONSTRUCTION PROJECTS

Committee:

Calin M. Popescu, Supervisor

Richard L. Tucker

John D. Borcherding

John D. Borcherding

James T. O'Connor

William R. Kelly

PARAMETRIC COST ESTIMATING MODEL FOR CONCEPTUAL COST ESTIMATING OF BUILDING CONSTRUCTION PROJECTS

by

KAN PHAOBUNJONG, B.S., M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas at Austin
May, 2002

สำหรับเจ๊ก แม่ เฮียหยู น้องนิน และริ่

Acknowledgements

This research effort would not have been possible without the help and encouragement of the many individuals who have been there for me along the way to completing this dissertation. First, I would like to express my appreciation to the members of my advisory committee for their direction and guidance. Special thanks go to my supervising professor Dr. Calin M. Popescu, for his invaluable guidance, support, and understanding throughout this research effort. I also want to thank all the other members of my supervising committee: Dr. Richard L. Tucker, Dr. John D. Borcherding, Dr. James T. O'Connor, and Dr. William R. Kelly for their advice, insightful guidance and support. Special thanks are extended to Dr. Richard L. Tucker for his utmost consideration and generosity in assisting with the funding during for my final year of study.

Special thanks are also extended to Dr. Schiller Liao at the Office of Facilities Planning and Construction (OFPC) at the University of Texas at Austin for his generous time and valuable advice. I also want to thank Mr. Jeff Rogers and Mr. Eric Short of OFPC for their time and assistance with the information gathering and data collection. Thanks are also due to Dr. Roger Elliot, Mr. William Beckham, and Mr. Keith Weldes at the Texas Higher Education Coordinating Board (THECB) for their generous time and assistance throughout the data collection at THECB. Without their support, this research could not have been successfully accomplished.

I also would like to thank Dr. Walter Meyer and Yim Ping Meyer for their continuous encouragement, care and concern for me, and for the many wonderful dinners and many loaves of bread. Their hospitality makes me feel truly at home in Austin. Special thanks are also extended to Noppadon Kamolvilassatian for his generous time in assisting with the computer program development. I also would like to thanks my colleagues and friends in the CEPM program, Santi Chinanuwatwong, Nuntapong Ovararin, Unsuk Jung, Marvin Oey, Walter Fagerlund, Hyoungkwan Kim, and Eunhee Kim for their friendship, encouragement and many meaningful discussions during my time in the program.

Finally, I wish to thank my beloved parents and brothers for their never-ending love, support and encouragement, and of course my Ree for her patience and understanding during all the years of my study. I owe everything to them.

Austin, Texas   Kan Phaobunjong
May 2002

PARAMETRIC COST ESTIMATING MODEL FOR CONCEPTUAL COST ESTIMATING OF BUILDING CONSTRUCTION PROJECTS

Publication No._____________

Kan Phaobunjong, Ph.D.
The University of Texas at Austin, 2002

Supervisor: Calin M. Popescu

It is well known that decisions at early stages of a construction project have the highest impact on the subsequent project performances. One of the bases for these important decisions is the project's conceptual cost estimate. It is the first serious effort at attempting to predict the cost of the project. Conceptual estimates are made in the early phases of a project before construction drawings are completed, often before they are begun. The primary function of a conceptual estimate is to tell the owner about the anticipated cost, thus being useful information for the owner in contemplating the project feasibility and further project development.

This research identifies building parameters that significantly influence the cost of building construction projects. The research presents methodologies for data collection, database development, data analyses, and parametric cost estimating model development for the purpose of performing building conceptual cost estimate. The data used in the model development and validation were based on the historical project data collected from the Texas Higher Education Coordinating Board (THECB). The data consist of 168 new building projects, built from 1990 to 2000. The technique of multiple regression analysis is used to develop the parametric cost estimating model by establishing the cost estimating relationships between the building parameters and the building construction cost.

The purpose of this study is to develop a reasonably accurate and practical method of systematized conceptual cost estimating that can be used by organizations involved in the planning and execution of construction projects, such as THECB and the offices of facilities planning of the major universities in Texas. The concepts and methodologies presented in the study can also be readily applied to other similar organizations.

Table of Contents

List of Tables

List of Figures

CHAPTER 1
INTRODUCTION

1.1 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 the 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 costs 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 factors affecting the project costs are unknown. Future 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 the conceptual cost estimate processes, especially for building construction projects. Estimating building construction costs can be difficult as most building projects are unique. A building construction project is a very complex undertaking, which can be composed of hundreds or thousands of construction activities called 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 building construction processes and ultimately their costs.

The complexity of the building 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 [Bley 1990]. This concept can be seen in Figure 1.1. Underestimates mean 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, 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.

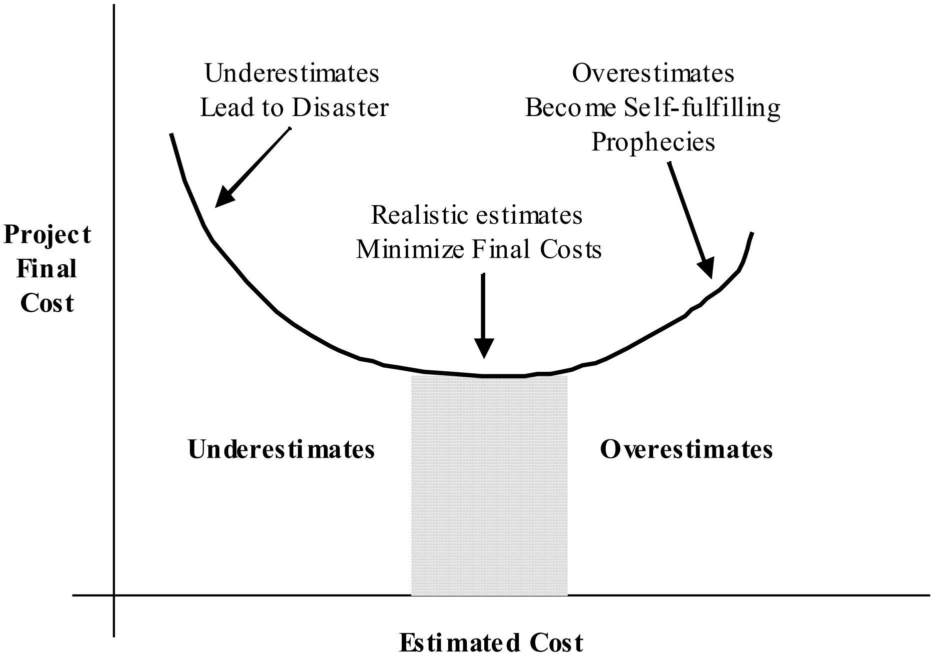


Figure 1.1. The Freiman Curve.

(Adapted from [Daschbach et al. 1988] and [Bley 1990])

Current practices in building project's conceptual cost estimating 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 require 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, will require a lot of resources for its development and implementation, such that only large organizations can afford.

In addition, construction cost estimating is generally more of an organization problem 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 better 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.

The limited research in the area of conceptual cost estimating, especially for the building construction industry, and the need for a better conceptual cost estimating methodology and tools are the motivation for this study.

1.2 DESCRIPTION OF THE PROBLEM

Literature reviews highlighted the need for estimating methodologies that utilize relevant historical project data for the estimates' development, yet not much research is carried out on this topic. Current practices of conceptual cost estimating have been performed mainly through the utilization of published cost files. 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 building construction projects. This is primarily because of the limited knowledge on the importance and potential applications of these historical data. Consequently, there has been little research to address the relationships of project characteristics to the project costs.

In addition, there is also a lack of knowledge on what specific data to collect, how to collect them, and what can be done to 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 project. There is a need to identify the information to be recorded and how they should be recorded so as to facilitate data collection and data consistency. Only in this way can databases be developed, and future analysis can be performed on the data and quantitative cost estimating model developed as a tool to assist in the cost estimating and analysis processes.

The cost of a building is a function of many variables. The first and basic problem is the identification and selection of these variables that may be used to describe a project and define its cost. Such variables must be measurable for each new building project. The second problem is the determination of the cost estimating relationships (CERs) function in terms of the selected variables in closed mathematical form. The criterion for the selection of variables and the technique chosen to construct the CERs function is dictated by the availability of data on such variables from building projects completed in the past. In addition, these data are also required for solving the third problem, i.e. testing the reliability of the derived function.

1.3 PROBLEM STATEMENT

Conceptual cost estimating in a building construction project is a difficult and generally a subjective process with no set standard 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 developed.

1.4 RESEARCH OBJECTIVES

The main objective of this research is to develop an accurate and practical method of systematic conceptual cost estimating that can be used by organizations involved in the planning and execution of building construction projects.

From this main research objective the following three sub-objectives are also identified:

1. To develop a parametric cost estimating model for conceptual cost estimating of building construction projects. The model development process is illustrated and discussed to promote the understanding of the model development requirements, methodologies used, and specific development outcomes.

2. To identify and assess the relative importance of the significant building characteristics or parameters to be incorporated into a cost estimating model to improve the model's cost estimating performance in the early phase of the project development, prior to 30% design completion.

3. To develop a computer program, based on the developed parametric cost estimating model, as a tool for performing the conceptual cost estimating.

1.5 RESEARCH SCOPE AND LIMITATIONS

The scope of this research is limited to the development of a parametric cost estimating model for conceptual cost estimating of building construction projects. The research is limited to the investigation and analysis of new building construction projects, related to state owned education facilities built by the various public institutions in Texas from 1990 to 2000. Examples of these public institutions are the University of Texas, the University of Houston, and Texas A&M University. The full list of all the 46 institutions can be seen in Table 4.1.

Due to the exploratory nature of this research, the focus of the research is only on the building construction costs of the project and does not include the analysis of other costs associated with fixed and movable equipment, engineering and design, and construction contingencies. In addition, site work costs will also not be included in the analysis due to the inherent dependency on the new site conditions, which must be considered separately from the cost model developed.

1.6 DISSERTATION ORGANIZATION

This dissertation is organized into seven chapters. It also includes four appendices containing supporting information used in the study, and the developed computer program codes. Chapter 2 presents the research background. It provides a summary of a very comprehensive literature review that focuses on conceptual cost estimating. Chapter 3 discusses the research methodology used in this study. It describes the research procedures employed, highlighting the processes involved and the intermediate results attained. Chapter 4 presents a brief overview of the data collected for this research. The details of the data are presented along with summary statistics. Chapter 5 focuses on the data analysis performed in this research. The analysis tools, methodologies and steps are presented along with the results. The development of the parametric cost estimating model through the use of multiple regression analysis technique is clearly presented. The validation of the model is also presented along with the conclusions and limitations observed for these data analysis steps. Chapter 6 discusses the computerized approach to conceptual cost estimating and presents the Parametric Cost Estimating Model (PCEM) computer program. PCEM computer program is developed to demonstrate the implementation of parametric method to conceptual cost estimating. A summary, conclusions, and recommendations for additional research are discussed in Chapter 7.

CHAPTER 2
RESEARCH BACKGROUND

2.1 INTRODUCTION

Project cost control is one of the primary duties of a project manager. Cost management is concerned primarily with the cost of the resources needed to complete project activities. When integrated with scope/quality management and time management, the three functions form the core of project management [PMI 1996]. In listing the reasons for a project success, the management of costs is one of the most important considerations of all project management. The main issue that is of concern for most owners is the project's "bottom line," or total cost. Cost management is essential to all phases of the construction project, from conceptual through engineering and design, and execution. Figure 2.1 illustrates the three general project phases in a typical construction project.

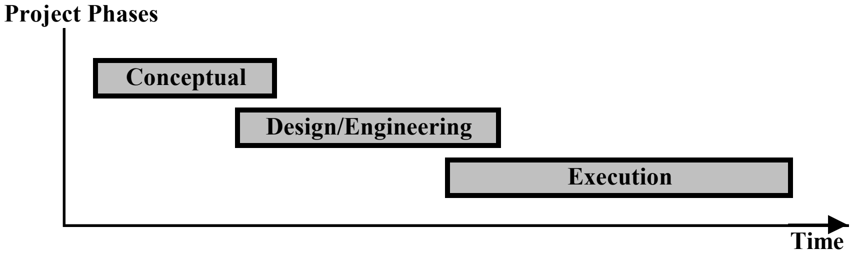


Figure 2.1. Construction Project Phases.

Cost management consists of two basic subtasks of cost planning and cost control. The process of planning and control changes as the project progresses along the project phases. In the conceptual phase, one of the activities of cost planning involves the preparation of the project conceptual cost estimate. The cost control activity in this phase is the establishment of the total project budget. The output from this conceptual stage then becomes the basis for planning and control of the following design and engineering phase. As design and engineering progresses, more and more elements of the project are identified and defined, providing more information for the subsequent revised cost estimate. This more detailed estimate then becomes the basis for the procurement and construction of a building, and subsequently the basis for controlling project cost in that execution stage. Figure 2.2 shows the simplified cost planning and control processes in a construction project.

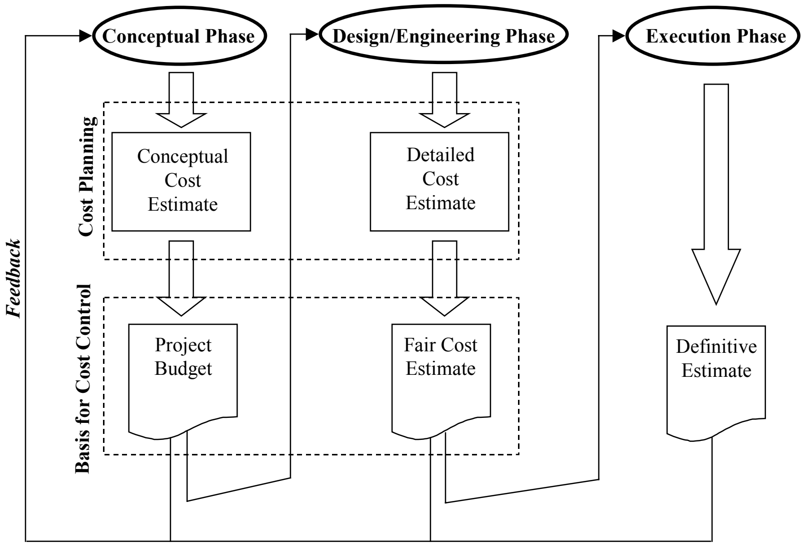


Figure 2.2. Project Cost Planning and Control Process.

Different phases in project development present different problems to the cost planning and control processes. The cost planning and control processes in all the phases are related, the output of the previous stages becoming the basis for planning and control of the subsequent stages. It also becomes evident that conceptual cost estimate is one of the first cost planning outputs in a project.

This chapter presents a comprehensive literature review and discussion of the principle concepts of conceptual cost estimating. Section 2.2 discusses the different types of estimates used in building construction projects. Section 2.3 focuses on conceptual cost estimating basics. Section 2.4 presents the input elements to conceptual cost estimating. Section 2.5 discusses the parametric method applications in conceptual cost estimating of a building construction project. In closing, the importance of conceptual cost estimating is discussed in Section 2.6.

2.2 TYPES OF COST ESTIMATES IN BUILDING CONSTRUCTION PROJECTS

Before the in-depth discussion of conceptual cost estimating, it is important to discuss the general types of cost estimating practices in building construction projects. Cost estimating involves developing an approximation (estimate) of the costs of the resources needed to complete project activities [PMI 1996]. 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 an assessment based on specific facts and assumptions of the final cost of a project, program, or process [Uppal 2000]. As illustrated in the previous section, 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.

There are many types of cost estimates that can be performed on a project, each type having different levels of accuracy. The estimating process becomes increasingly more expensive as more detailed and accurate techniques are applied [Barrie and Paulson 1992]. Analysis of different classifications of estimates concludes that there are two main types of estimates: conceptual and detailed [AACE 97, Barrie and Paulson 1992, Bley 1990, PMI 1996, Popham 1996].

2.2.1 Conceptual Estimate

A conceptual estimate is also known as a top-down, order of magnitude, ballpark, 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%.

A conceptual estimate is a pre-design 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 estimate derivation methodologies are usually those relying on historical information to predict future cost of the new project, such as referencing previously completed project to estimate the cost of a new project that is similar in nature.

2.2.2 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 indicate clearly the required quality of materials and workmanship. A detailed estimate is prepared from very 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.

2.2.3 Other Types of Estimates

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 [Bley 1990].

The various estimates discussed above are carried out in sequence, the previous cost estimate being the input to the next one. The estimates are successively refined, incorporating new information and thus keeping a continuously updated estimate that becomes the budget, available for control process [Barrie and Paulson 1992]. 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.

2.2.4 AACE Cost Estimate Classification System

In 1997, the American Association of Cost Engineering (AACE) produced a Recommended Practice No. 17R-97: Cost Estimate Classification System. The AACE guideline is developed with the intention of providing a generic methodology for classifying project cost estimates in any industry. The purpose of the guideline is to provide a common understanding of the concepts involved with classifying project cost estimates. The guideline establishes the quality or level of information needed for the various estimate preparations, the end usage of various estimate, the methods used in preparing the various estimates, the accuracy levels expected from the various estimates, and the preparation effort required for the various estimates. By providing a means for classify an estimate and define its accuracies and the required effort, the guideline helps to promote communication among all project participants involved with project estimate preparation, evaluation and application.

AACE Recommended Practice No. 17R-97 suggests that the most significant factors for classifying the cost estimate are the level of project definition, end usage of the estimate, estimating methodology, and the effort and time needed to prepare the estimate. The guideline uses the level of project definition as the “primary” characteristic for classifying cost estimate. The other factors are considered the “secondary” characteristics of the estimate. Figure 2.3 illustrates the five classes of cost estimate as proposed by AACE, along with their primary and secondary characteristics.

	Primary Characteristic	**Secondary Characteristic**			
Estimate Class	**Level of Project Definition** Expressed as % of complete definition	**End Usage** Typical purpose of estimate	**Methodology** Typical estimating method	**Expected Accuracy Range** Typical variation in low and high ranges [a]	**Preparation Effort** Typical degree of effort relative to least cost index of 1 [b]
Class 5	0% to 2%	Screening or Feasibility	Stochastic or Judgment	L: -20% to -50% H: +30% to +100%	1
Class 4	1% to 15%	Concept Study or Feasibility	Primarily Stochastic	L: -15% to -30% H: +20% to +50%	2 to 4
Class 3	10% to 40%	Budget, Authorization, or Control	Mixed, but Primarily Stochastic	L: -10% to -20% H: +10% to +30%	3 to 10
Class 2	30% to 70%	Control or Bid/Tender	Primarily Deterministic	L: -5% to -15% H: +5% to +20%	5 to 20
Class 1	50% to 100%	Check Estimate or Bid/Tender	Deterministic	L: -3% to -10% H: +3% to +15%	10 to 100

Notes: [a] The availability of applicable reference cost data can affect the range markedly.
[b] If the range index value of "1" represents 0.005% of project costs, then an index value of 100 represents 0.5%. Estimate preparation is highly dependent upon the size of the project and the quality of the estimating data.

Figure 2.3. AACE Generic Cost Estimating Classification Matrix.

(Adapted from [AACE 1997] and [Lorance et al. 1999])

As can be seen in Figure 2.3, the five estimating classes are labeled from 1 to 5, with Class 1 being the most accurate estimate requiring the most effort, while Class 5 is the least accurate and requiring the least amount of effort to prepare. The level of project information required increases as the estimate changes from Class 5 to Class 1. The methodology used also changes from judgmental* or stochastic** for Class 5 estimate to a completely deterministic*** method for Class 1 level estimate. This is because at Class 5, the level of information available is limited, thereby lending itself to only judgmental or stochastic estimating process. As the level of project definition increases, the estimating methodology moves from the stochastic process to the deterministic process.

---

* judgmental process refers to a non-quantitative process whereby conclusions and decisions are made based on the experiences and opinions of the decision makers.

** stochastic process, also refers to as probabilistic process, is a process whereby at least one of the input to the stochastic model is uncertain and subject to variation.

*** deterministic process refers to the process where by all the inputs to the deterministic model are known and cannot vary.

Based on the level of definition, end usage, and the methodology used, Class 3, 4, and 5 may be considered as the subgroup under the conceptual cost estimate category as described in Section 2.2.1, while Class 2 and 3 may be considered under the budget estimate category, and Class 1 and 2 may be considered under the detailed estimate category.

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

2.3 CONCEPTUAL COST ESTIMATING BASICS

Conceptual cost estimating is an important pre-design planning process. This section provides thorough discussions on the conceptual cost estimating basics. The following subsections present the conceptual cost estimating definitions, characteristics, processes, and outputs.

2.3.1 Conceptual Cost Estimating Definition

A "conceptual estimate" is an estimate prepared by using engineering concepts and avoiding the counting of individual pieces [Kreps et al. 1990]. As the name implies, conceptual estimates are generally made in the early phases of a project, before construction drawings are completed, often before they are hardly begun. The first function of a conceptual estimate is to tell the owner about the anticipated cost, thus presenting useful information for the owner in contemplating the project feasibility and further development [Barrie and Paulson 1992]. A conceptual estimate is also used to set a preliminary construction budget, and to control construction costs at the most critical stage, during the design. Bley defined conceptual cost estimating as the forecast of project costs that is performed before any significant amount of information is available from detailed design and with a still incomplete work scope definition, with the purpose of using it as the basis for important project decisions like go/no go and the appropriation of funds decisions [Bley 1990].

2.3.2 Conceptual Cost Estimating Characteristics

The first recognized characteristic of conceptual estimating, like all other estimating, is the inexactness in the process. Estimating is defined by Webster's Dictionary as "to form an approximate judgment or opinion regarding value, or to calculate approximately." Estimating is not an exact science and involves many subjective judgments to qualify and quantify many of the variables that can impact the construction cost. Bley states that conceptual estimating is an inexact process based to a large degree on judgment and experience due to the lack of information and always existing uncertainty at the conceptual stage [Bley 1990]. Noble has stated that an estimate is a judgment, opinion, forecast, or prediction of cost, time, or other project aspects [Noble 1987]. Ostwald stated that personal opinion is inescapable in estimating. In addition he also added that it may be easy to be critical of opinion, but in the absence of data and with shortage of time, there may be no other way to evaluate designs but to use opinion [Ostwald 2001]. Skidmore expressed that conceptual estimating is a rather imprecise business due to the high level of complexity and uncertainty found in the construction industry [Skidmore 1986]. Conceptual estimating is a mixture of art and science; the science of estimating tells the cost of past work. The art is in visualizing a project and the construction of each detail, selecting comparative costs from past projects and adjusting them to new conditions [Carr 1983].

The second characteristic of conceptual estimating is that its accuracy and validity are highly related to the level of information provided by the project scope. An estimate is an evaluation of all the costs of a project under consideration as defined by an agreed-upon scope [Ludwig 1988]. Noble suggests that estimating is the art of approximating the probable cost of a project based on the information available at the time [Noble 1987]. In addition, Bley states that the availability of a good, complete scope definition is considered the most crucial factor for conceptual estimating [Bley 1990]. Finally, Di Natale has given a more general specification by expressing that conceptual estimates are in reality a monetary proposal of a tentative design [Di Natale 1980].

The third characteristic of conceptual estimating is that it is a resource restricted activity. The main resources for conceptual estimating are information, time, and cost. Due to the fact that conceptual estimating is performed in the early stages of the project, the scope information available is usually restricted in detail as well as in precision [Bley 1990]. Logcher states that an estimate is a prediction of a future outcome based on incomplete and imprecise data; it is based on current knowledge of the project and current or past knowledge of costs [Logcher 1980]. In addition, the time and cost available for making the estimate is restricted. Conceptual estimating is used to determine the feasibility of a project quickly or screen several alternative designs [Popham 1996]. Therefore, the estimate, although important, cannot be given much time and resources. Conceptual estimating must be inexpensive, quick, and reasonably accurate [Kouskoulas et al. 1974].

2.3.3 Conceptual Cost Estimating Process

A generic conceptual cost estimating process is displayed graphically in Figure 2.4. The process begins with a request made by management to estimate the cost of a new project. The most important part of the request is the project scope [Bley 1990]. The first task for the estimator is to study and interpret the project scope and produce an estimating plan. This stage also involves the formulation of estimating methodology and determination of additional information needed. The next task is to collect additional information that is required for the conceptual cost estimate. This additional information can be separated into two categories: current data and historical data. Current data include those relating to current costs, productivities, project conditions and future trends. Historical data include those relating to past projects that are similar to the ones that are to be estimated. The selection and usage of these data is a critical part of the estimating process because the selection of inappropriate information will negatively affect the end quality of the estimate [Bley 1990]. The estimators must, at this point, conceptualize the project, understand the necessary activities to be done and when to be done, and compare the project with previous ones if possible [O'Connor 1985]. With all the information assembled, the conceptual cost estimating process can be performed. This stage is the heart of the estimating process. The outputs from this stage are the project conceptual cost estimate and a documented estimating basis used to develop this cost. It is very important to describe in detail all the information, assumptions, adjustments, and procedures considered in the estimate, as this document is necessary to support the estimate, as well as for future cost evaluation due to changes and modifications of project scope [Bley 1990]. The resulting conceptual cost estimate is then submitted to management for decision-making. At this point, the estimating process is finished in this cycle. The process may be reiterated so as to accommodate for changes and modifications to the project scope.

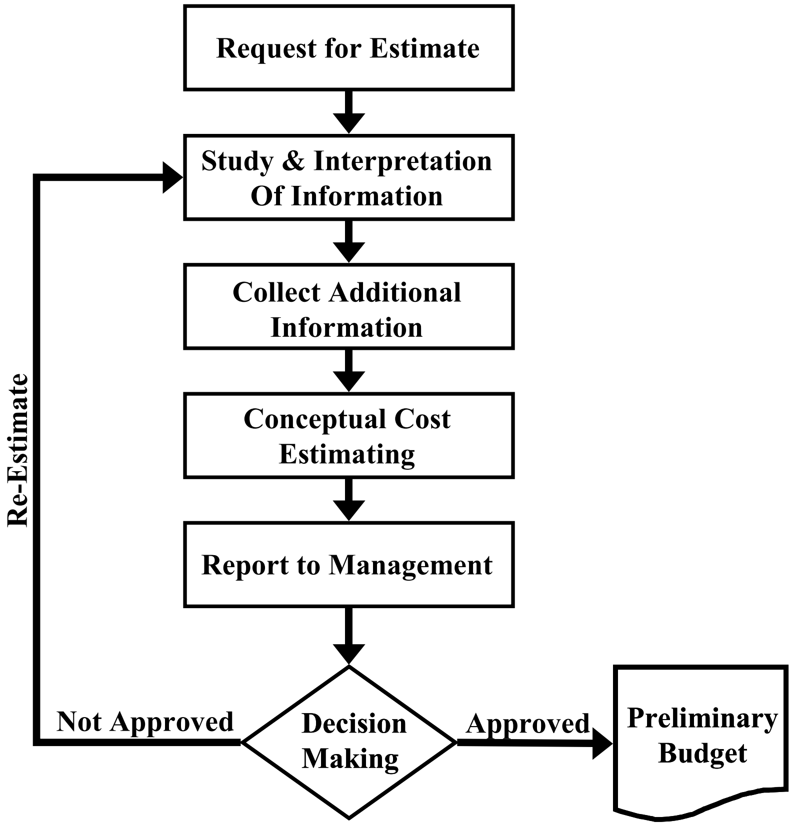


Figure 2.4. Conceptual Cost Estimating Process.

2.3.4 Conceptual Cost Estimating Output

The primary output of the cost estimating effort is the cost estimate. The estimate is typically expressed in units of currency. Alternative units can be work quantities, material quantities, or staff work hours. However, for majority of the construction projects, the units of currency are mostly applicable; therefore, they are frequently used. In this way, it also facilitates project evaluation and comparisons with other projects [PMI 1996].

In addition to the cost estimate, supporting detail for the estimates should be included as part of the estimate output. The supporting detail should include [PMI 1996]:

a. The description of the scope of work being estimated;
b. Documentation of the basis for the estimate;
c. Documentation of any assumptions made; and
d. An indication of the range of possible results, such as 50,000 ± 2,000 or between 48,000 and 52,000.

These supporting details not only provide a backup or explanation for estimate, but also enhance the estimate by providing more information about the estimate. With more detailed information, the estimate can provide more insight for the decision makers, allowing for the improvement in the decision-making process.

With the above definitions of conceptual cost estimating and the discussions on its characteristics, processes, and outputs, the concepts and meanings of conceptual cost estimating is established. The next section discusses the input elements to conceptual cost estimating.

2.4 INPUT ELEMENTS TO CONCEPTUAL COST ESTIMATING

Figure 2.5 shows the combination of all the input elements for successful conceptual cost estimating.

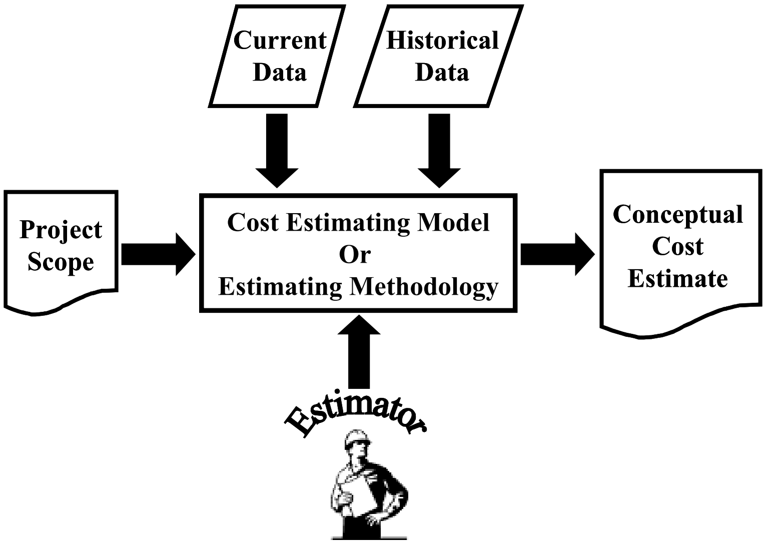


Figure 2.5. Conceptual Cost Estimating Elements.

From the discussion on the estimating process, the four elements required for conceptual estimating are identified as seen in Figure 2.5.

2.4.1 Project Scope

The project scope is the most critical element in conceptual cost estimating [Bley 1990, Cho 2000, Gibson et al. 1993, Smith et al. 1983]. Project scope definition is a process by which projects are defined and prepared for execution [Cho 1999]. The project scope must be prepared by owner and detailed by selected designer. During this process, information relating to general project requirements, essential materials and equipment, and construction methods or procedures are identified. This information is compiled into a project scope definition package. The package should have all the necessary information that permits the subsequent detailed design and engineering processes [Gibson et al. 1993]. Project scope is the main input of all cost estimating efforts. The conceptual scope definition effort attempts to resolve the following issues [CII 1995]:

a. Provide means to reach agreements on items to be included;
b. Provide data for generating usage requirement and further project development; and
c. Reduce uncertainty through project specifications and definition.

The project scope describes the nature of the project, setting the required components and specifications of the project. These then become the basis for the conceptual cost estimating process.

2.4.2 Information

Information is one of the principal resources for estimating [Stewart 1987]. Information gives the estimator a reference from where to derive cost figures [Bley 1990]. Two types of information are required for conceptual cost estimating: historical and current information. Historical information or historical data are data associated with previously completed projects, such as project specifications and costs. Most companies rely on their own historical data instead of using published data [Bley 1990]. Historical information, regarding what types of resources were required for similar work on previous projects, should be used if available [PMI 1996]. Current information or current data are the second element of the conceptual cost estimating information. As the name implies, it relates to the up-to-date data associated with the new project, such as current applicable local prices and indexes, productivities, and specific site conditions. The main sources for current information are usually published data. The historical data are the basis used to determine the expected cost of the new project based on completed similar past projects. The current data are used to adjust and fine-tune the estimated cost to fit the current project's conditions. Of the two, historical data are the most relevant and critical one [Bley 1990].

2.4.3 Estimating Methodology

The next 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. An office building construction project will be estimated differently than an industrial process plant construction project. Obviously, the nature of the project is different but more importantly, the cost components associated with the two projects are different. In this way, the two construction projects demand two different approaches to estimating the project construction costs. In addition, as can be seen in the previous discussions, conceptual cost estimating can span across three classes of estimates (Class 3 to Class 5) as prescribed by AACE. It encompasses all cost estimates made in the early phases of a project. These estimates have varying levels of information available when the estimates are 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 [Barrie and Paulson 1992]. Most of the existing conceptual cost estimating methods for building construction project fall into one of the following categories [Barrie and Paulson 1992, Bley 1990, Ostwald 2001, PMI 1996]:

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

The three methods are presented in the order of increasing methodology complexity, information required, effort required and estimate accuracy. The three 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.

2.4.3.1 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 [PMI 1996]. 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 [PMI 1996]. Application of analogous estimating is generally limited to AACE's Class 5 estimate.

2.4.3.2 Unit Method and Parametric 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 [Bley 1990]. 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. [Barrie and Paulson 1992, Karshenas 1984, Ostwald 2001, Parker 1984]. 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 [Bley 1990, Ostwald 2001]. Unlike the parametric method, the unit method generally refers to the estimating method whereby the estimates are determined from predetermined unit costs.

Similar to unit, parameters are the measures of the basic elements of the project [Orczyk et al. 1990]. The parametric method involves using project characteristic in a mathematical model to predict project costs. The model is developed by first establishing relationships between the building parameters and the building cost. Several methods can be used to correlate the historical costs and the parameter information, but the technique most commonly used is linear regression, although other mathematical functions have been successfully applied [Bley 1990, Bowlby et al. 1986, Karshenas 1984, Kouskoulas et al. 1974]. The model may be simple, having only one parameter, or complex, with multiple parameters.

More discussion on the applications of parametric estimating methods for conceptual cost estimating of building construction project can be seen in Section 2.5.

2.4.3.3 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 [Melin 1994]. 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.

2.4.4 An Estimator

The last and final element is the estimator. The estimator is the person responsible for organizing and analyzing all the information and finally putting together the estimate. The estimator is not only the coordinator of the process, but has also brought to the process his expertise and experience in cost estimating. Expert judgment will often be required to assess the inputs to this process [PMI 1996]. Expertise has been shown to be a necessary ingredient for conceptual cost estimating [Bley 1990]. Expertise means possessing the ability, competency, experience, familiarity, knowledge, skills, and mastery in a specific field. Expertise in cost estimating allows estimators to better understand a new project based on previously executed projects, generate information that is not available by making sound assumptions, and making adjustments to existing information based on current conditions. The estimator's ability to visualize the scope of work from incomplete scope definition is one of the most important factors in conceptual cost estimating [Bley 1990].

2.5 PARAMETRIC METHOD APPLICATIONS IN CONCEPTUAL COST ESTIMATING OF BUILDING CONSTRUCTION PROJECT

Bley states that parametric estimating is one of the most used methods for conceptual cost estimating [Bley 1990]. Parametric estimating has been applied extensively across various industries. Black uses a parametric estimating methodology for the preliminary cost estimating of a jet steam ejector. His model was based on deriving a mathematical function that uses the ejector's capacity, suction pressure, and steam consumption parameters for estimating the jet steam ejector cost [Black 1984]. Garza produced a parametric model whereby material cost per 100 ft of carbon steel pipe can be estimated from the following design parameters: pipe diameter, number of elbows, and flange rating [Garza et al. 1995]. Bode developed a parametric estimating model to estimate ball bearing manufacturing cost using the product design specifications: the inner and outer diameter, width, and design type [Bode 1998]. Creese applied the parametric estimating method to the cost estimation of timber bridges [Creese et al. 1995]. Al Khalil utilized the parametric method for determining the total project cost for reservoir construction based on the reservoir's capacity, construction duration, and access index parameter [Al Khalil et al. 1999]. Parametric method has also been used in the estimating of highway preliminary construction cost [Hegazy et al. 1998, Al Tabtabai et al. 1999].

Section 2.4.3.2 discusses briefly the parametric method as used in conceptual cost estimating of building construction project. To gain a better understanding of the parametric cost estimating method, this section focuses more on the topic of parametric estimating and its application in the building construction projects. The following subsections present its brief history, execution procedures, and discussion on applicable parameters. Lastly, the parametric estimating issues relating to the implementation for conceptual cost estimating of building construction projects is also presented.

2.5.1 History of Parametric Estimating

Before the development of parametric estimating, the quantitative method in conceptual cost estimating had been limited to performing the detailed quantity takeoff and pricing [Orczyk 1990]. Quantity takeoff involved a lot of data identification, computation, and documentation. It is a very demanding process. In addition, since the scope of the project is often unclear at the time of the conceptual cost estimate, many assumptions were required and used in order to perform the quantity takeoff and pricing. Consequently, the accuracy of these estimates depended upon the validity of the assumptions made.

In the early 1960s, the U.S. Air Force, was dissatisfied with this bottom-up estimating, so they asked the RAND Corporation to look for a better way of making program cost estimates. RAND developed the idea of a parametric estimating model whereby estimating relationships were found between project or program costs and one or more technical parameters of the project or program. This concept of cost estimating relationships (CERs) was presented to the Air Force, and the RAND study suggested that one or more technical parameters might be useful for predicting project or program costs [Black 1984, Krieg 1979, Orczyk 1990].

2.5.2 Parametric Estimating Procedure

The development of the parametric cost estimating model generally involves the following six steps. These steps are illustrated in Figure 2.6.

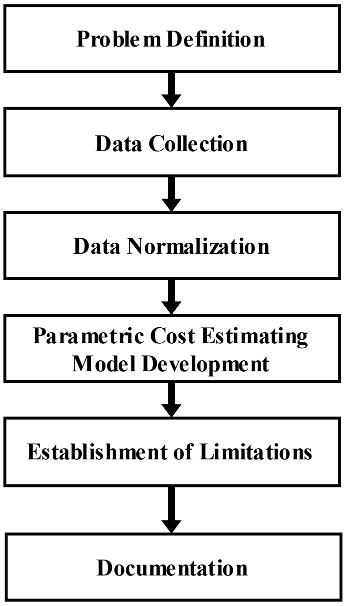


Figure 2.6. Parametric Cost Estimating Procedure.

(Adapted from [Black 1984])

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 the research study [Black 1984].

The second step is the data collection. Parametric estimating requires an extensive database, where historical records are extremely important [Rose 1982]. The design or engineering parameters that drive parametric cost estimates are developed from historical cost databases [Meyer et al. 1999]. 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 building construction project, the cost data for each project must be adjusted for time and location differences [Orczyk 1990]. This step is important and must be performed before further analysis can be made to the data.

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 [Ostwald 2001]. 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 [Black 1984, Orczyk 1990].

Recently, there have been numerous experiments of neural network methods as an alternative to the statistical method of developing the parametric cost estimating model. Neural networks have evolved based on artificial intelligence and is commonly used for difficult tasks involving pattern detection and recognition that elude conventional analytical techniques [Hegazy et al. 1998]. Although recent research demonstrated the potential use of this technique in construction [Moselhi et al. 1992], the neural network method suffered from the lack of transparency in the derivation process. It is regarded as the black box approach, with neural processing difficult to be traced and explained [Hegazy et al. 1998]. The methodology also requires trial and error in setting the connection weights during the model development [Hegazy et al. 1994]. Bode concludes that the neural network research to date is quite experimental because neural network theory does not yet provide applicable rules, optimal setting of control variables, and topologies [Bode 1999].

Krieg states that the success of the parametric model depends on capturing the repeatable characteristics of the building projects to be estimated [Krieg 79]. The objective of parametric cost estimating is to produce a set of algorithms or mathematical expressions that reflect the world that is observed using a minimum of data.

The fifth step of the parametric estimating procedure is the establishment of model limitations. The model is generally developed from a limited data set, is therefore valid only for the ranges of variables used in the model [Orczyk 1990]. 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 [Rose 1982]. Extrapolation beyond this range will produce unverified results [Orczyk 1990].

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 [Black 1984]. Information and meanings of the terms used in the data collection and the model development must be documented along with all the calculation methods.

2.5.3 Parameters for Building Construction Project

Many papers and textbooks have suggested the relationships between individual building parameters and the cost of the building components. Table 2.1 presents the parameters identified and the related cost components.

Table 2.1. Building Parameters and the Related Building Cost Components.

Building Parameter	**Related Cost Components of Building**
Source: Dell'Isola and Kirk 1981	
1. Footprint Area at Grade	a. Foundation b. Substructure
2. Area of Suspended Floors	a. Superstructure-floor
3. Area of Roof	a. Superstructure-roof b. Roofing
4. Area of Exterior Wall	Exterior Cladding
5. Area, Exterior Doors/Windows	Exterior Doors/Windows
6. Total Area Finished	Interior Construction
7. Gross Site Area	Site Work
8. Total Enclosed Volume	Mechanical
9. Transformer Rating	Electrical
Source: Collier 1984	
1. Total Area of Building	a. Floors b. Roofs c. Partitions d. Mechanical e. Electrical
2. Perimeter of Building	a. Exterior Cladding b. Eaves & Parapets
3. Number of Occupants	a. Elevators b. Plumbing

(Adapted from [Orczyk 1990])

These parameters are used primarily in the development of the estimate based on the unit method. That is, the costs of the building components are estimated based on the proposed project quantities and the pre-established unit costs. For example the cost of the foundation can be calculated based on the footprint area at grade and the pre-established unit cost for the foundation, such as the foundation cost being $2.50 per square foot area of the building foot print. By determining the costs of each of these building components with the desired building quantities and the pre-established unit costs, the total estimated cost of the building is compiled.

Alternatively, extensive literature reviews have uncovered three studies of parametric method as applied to building construction project.

The first was by Kouskoulas and Koehn [Kouskoulas et al. 1974]. In this study, the parametric estimating method was developed to estimate the unit cost of the building, in dollars per square foot. Regression analysis was performed on the data collected from 38 buildings built between 1963 and 1972. The significant parameters identified were:

a. Region of U.S. where building was located
b. Year constructed
c. End use of building
d. Number of stories

e. Quality of workmanship and materials specified

The second study identified was by Karshenas [Karshenas 1984]. Karshenas used regression analysis to construct the parametric cost equation to estimate the total building cost. The study focused on the construction cost of a multi-story office buildings. The data used in the study consisted of 24 office buildings built between 1961 and 1979. Only two building parameters were used in the study and in developing the parametric cost estimating model. The two parameters were:

a. Typical floor area
b. Height of the building

The third study was by Bowlby and Schriver [Bowlby et al. 1986]. This study used regression analysis to develop a parametric estimating model for predicting the cost of buildings in dollars per square foot. Bowlby and Schriver had access to F.W. Dodge contract construction data, and consequently a total of 157,855 projects, executed between 1972 and 1982, were extracted from the database and used in this study. The study identified the following parameters as significant:

a. Number of stories
b. Total area of building

c. Metropolitan or rural location
d. Winter start in a northern state
e. End use of building
f. Building framing types
g. Region of U.S. where building is located

As can be seen in the above presentation, numerous building parameters are suggested by both studies. However, it is evident that the unit method requires data at a more detailed, micro level, than those data used by the parametric method. The data used in the parametric methods tends to be more general, at a macro level. This discrepancy is further discussed in the next section.

2.5.4 Parametric Cost Estimating Issues

Numerous building parameters are identified through many studies and research in conceptual cost estimating based on the unit method and the parametric method. The nature of the parameters identified is different because of the fundamental differences between two methods. That is, the unit method is a simpler approach in which the unit costs (parameter and cost relationships) are generally pre-established. In contrast, the parametric method makes no use of any prior or pre-determined relationships or rules of thumb. Through the parametric method, the significant parameters, which are also known as the cost drivers are identified, and the relationships between the parameters and cost are established based on the historical project data. In this manner the parametric method is a more data intensive process and demands sufficiently large amounts of historical project data to ensure statistical validity of the derived parametric model. To develop a project database with a sufficient numbers of projects and with data consistency means that the level of detail must be sacrificed. This explains why parameters used in the parametric method tend to be more high level and not as detailed as those used in the unit method.

The first issue identified relating to the parametric method is thus the historical data, or the lack of it. Park has stated that parametric estimates are scarcely used because no systemized historical parametric databases exist. Instead, cost estimates derived by unit method are more commonly used [Park et al. 1999]. This lack of data is among the reasons contributing to limited research and application of the parametric method in conceptual cost estimating of building construction project.

The second issue identified stems from limited research conducted on the parametric method. Limited research has lead to very little knowledge on the implementation of parametric method and limited identification of the parameters that can be successfully used in developing a parametric cost estimating model.

The third issue identified is that the parametric method has the potential of being the ideal methodology for conceptual cost estimating of building construction projects. Although a sufficient amount of data is required for its development, the parametric cost estimating model, once developed, provides the following desirable features:

a. Based on relevant historical projects, it reflects the organization's experiences and degree of specialization in construction projects involved;
b. Permits conceptual cost estimating with a few high level project parameters, facilitating the estimating process and leading to an efficient and quick method of producing the estimate; and
c. A systematic and quantitative approach minimizes the subjectivity in the estimating process and ensures consistency in the estimates generated.

2.6 IMPORTANCE OF CONCEPTUAL COST ESTIMATING

Conceptual cost estimating is an important function in project cost management. A conceptual cost estimate is vital to the following areas of project cost management: 1) early decision making, and 2) project cost planning and control.

2.6.1 Early Decision Making

Conceptual cost estimate is one of the most important pieces of information for decision making at the conceptual stage [Bley 1990]. It is well known that decisions at early stages of a construction project have the highest impact on the subsequent project performances [Barrie and Paulson 1992, CII 1995]. Figure 2.7 illustrates this concept. It shows that the level of influence on the final project cost decreases as the project progresses from conceptual into execution phase. The curve reflects the ability of management to affect the cost outcome of the project during the various stages of a project. According to Daschbach, decisions at the conceptual stage affect 70 to 80% of the total eventual costs even though engineering is only 1 or 2 % complete [Daschbach et al. 1988].

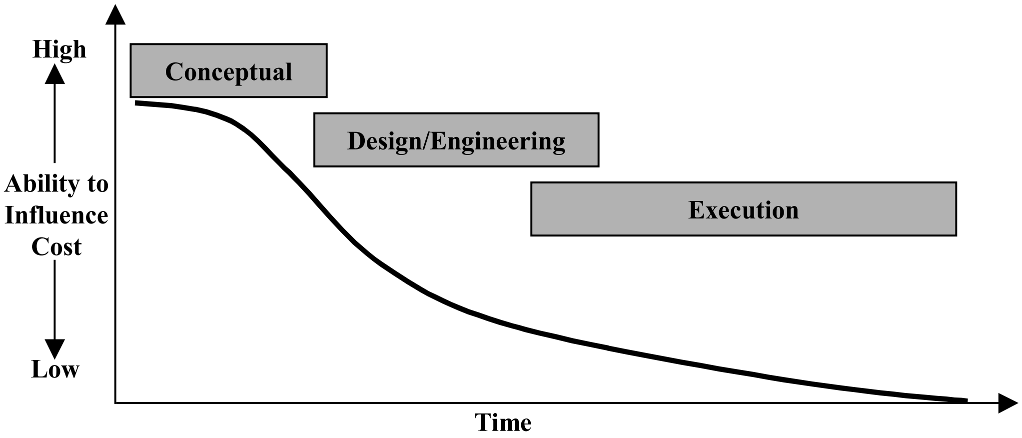


Figure 2.7. Influence Curve of Decisions on Project Cost Performance.

(Adapted from [CII 1995])

Conceptual cost estimates are the basis for important early project decisions. Project feasibility is often determined by the conceptual cost estimate [Barrie and Paulson 1992, Bley 1990, PMI 1996]. Finally, conceptual cost estimates are also critical in decisions related to modifications of the general scope and reevaluating of a project.

2.6.2 Project Cost Planning and Control

Conceptual cost estimating is part of the project planning function that forms the basis for assigning financial resources to the project and creating the project budget [Bley 1990]. With the budget, cost control can be exercised in the subsequent phases by ensuring that all work performed satisfy the budget requirement. The risk of a cost overrun of a project can be reduced by preparing an accurate budget before planning, funding, and design [Mitchell 1995]. In this sense, a conceptual cost estimate produces the first formal cost goal for a project.

2.7 CHAPTER SUMMARY

This chapter provides a comprehensive background for the research on conceptual cost estimating as applied to building construction projects. From the discussions on various types of cost estimate, a conceptual cost estimate is identified as the first serious attempt at predicting the cost of the project. A conceptual cost estimate is a pre-design estimate, usually performed with limited information. The four essential inputs to conceptual cost estimating are project scope, information, estimating methodology, and an estimator. One of the most used methods for conceptual cost estimating is the parametric method. A parametric method involves establishing the relationships, based on past experiences, between building characteristics or parameters and building costs. These relationships are then used to estimate the costs of future projects. Based upon this literature survey, the parametric method can benefit from additional research and may prove to be the ideal methodology for conceptual cost estimating of building construction projects. The chapter concludes by discussing the importance of conceptual cost estimating and the vital roles it plays in early decision making and the project planning and control effort.

The next chapter presents the research methodology employed in this study.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 INTRODUCTION

This research study is a retrospective study whereby past project data are collected and analyzed. The research areas addressed by this dissertation are the building construction project cost drivers identification and the establishment of relationships between these cost drivers and the cost characteristics of building construction projects. The research also investigates and demonstrates the data collection process, electronic database development, statistical analysis of data, and cost estimating model development, validation, and implementation.

This chapter presents the research methodology employed by the author in the development and execution of this research study. First, the overview of the research process is presented before each research step is discussed. The significant research steps include literature reviews, issue identification, data collection, data preparation, data analysis, Parametric Cost Estimating Model (PCEM) computer program development, and documentation. Figure 3.1 illustrates graphically the research flow chart.

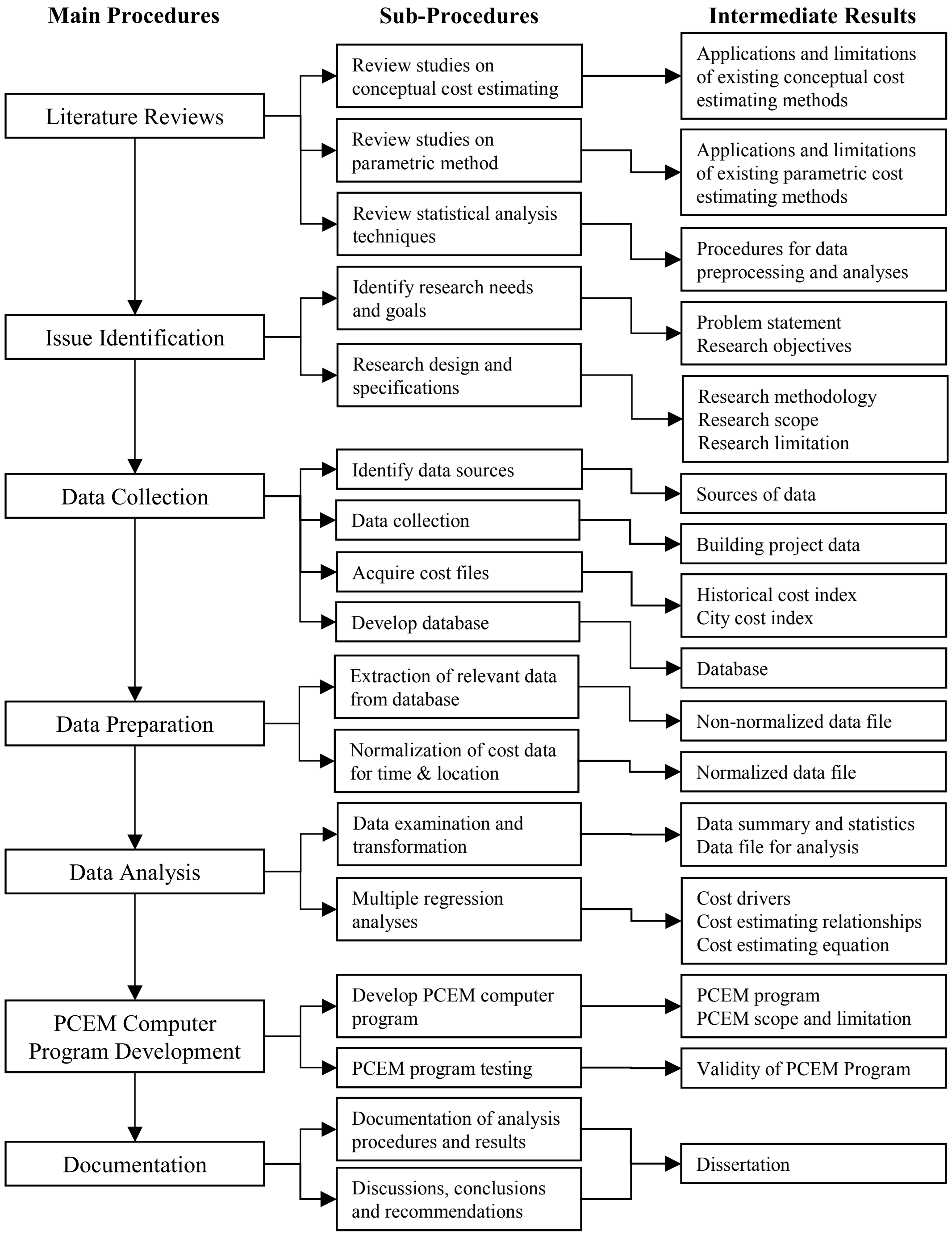


Figure 3.1. Research Flow Chart.

3.2 LITERATURE REVIEW

The literature review provides the background information or framework for the current research study. A literature review is a very important part of any research effort. It is important to discuss any previous related research study in the proposed area. In discussing this previous investigations, it can be demonstrated how the current research is related to the previous studies.

For this research, the literature review focuses on issues relating to conceptual cost estimating and parametric methods as applied to building construction projects. The review of conceptual cost estimating concepts and methodology is important in understanding the current processes and their needs. A parametric method is then reviewed so as to establish its current state-of-the-art and potential application in conceptual cost estimating of building construction costs.

Statistical analysis techniques are also reviewed as part of the preliminary research plan so as to develop the procedures and identify any requirements for the subsequent data preprocessing and analysis steps.

3.3 ISSUE IDENTIFICATION

From the preliminary investigations and literature review, the pertinent issues for this research study are highlighted and considered in the definition of the research study. The needs and goals of the research are first established. With respect to the recognized needs and the goals set, the research is designed and specified. Through this research definition stage, the specific problem statement, research objectives, methodology, scope, and limitations are identified and established. These items clearly define the research study, guiding the research effort, and form the basis for the subsequent research evaluations.

3.4 DATA COLLECTION

Data collection is defined as an activity of constructing primary data records for a given sample or population of observations [Knoke and Bohrnstedt 1994]. There are five different types of research with different modes of observation [Babbie 1992]. The five research types are listed and described as follows:

1. Experimental Research	Involves taking action and observing the consequences of that action.
2. Survey Research	Involves collecting data through asking people questions.

3. Field Research	Involves the direct observation of social phenomena in natural setting.
4. Unobtrusive Research	Involves investigation without the researcher intruding into whatever is being studied. It encompasses content analysis, analysis of existing statistics, or historical/comparative analysis.
5. Evaluation Research	Involves evaluation of the impact of social intervention by using experimental or quasi-experimental method.

Because of the nature of this study, and the types of data involved, only the unobtrusive research method applies. With this method, existing data is collected for use in the analysis. One of the disadvantages to this method is that the use of existing data, which is also referred to as secondary data, means that the researcher has less control over the data contents, quality and quantity. However, the principal advantage to this method is that it relies on less involvement and interactions with other parties as data are often retrieved from the available documents.

3.4.1 Types of Data Required

The data needed for this research can be separated into two parts: historical or past project data set and cost indexes data set.

3.4.1.1 Historical Project Data

The historical project data is the most crucial input to the parametric cost estimating model development. The nature of the data collected, such as the data contents, level of details, and number of cases not only dictate the selection of variables, or building parameters, to be analyzed, but also influence the analysis methodology. The data is not only required for the building parameters identification and the construction of the cost estimating function, but is also necessary for solving the third problem of testing the reliability of the derived cost estimating function.

3.4.1.2 Cost Indexes Data

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

The cost indexes data set for this research refers primarily to the city cost index and historical cost index. The dollar amount of past projects must be adjusted for the differences in locations and times associated with the projects. Only in this manner can the project costs be compared and correctly analyzed. In addition, the cost indexes are also needed later, for the reverse role of adjusting the cost estimate produced, which is at a reference base, to a specified location and time requirement of the new building that is being estimated.

3.4.2 Data Sources

The first step to the data collection process is to identify the sources of data. The collection and compilation of the building construction cost data is a challenging task due to the lack of good and consistent record keeping. Few organizations have the necessary know-how and resources to develop and maintain a good historical project database. Unfortunately, those organizations that do also consider such information to be proprietary information, adding values to the organization and enhancing the organization's competitiveness. The developed databases are usually kept and shared in-house. Therefore, for this research only public agencies were contacted for the data acquisitions.

With respect to the historical data set, three data sources are pursued in this research. The first is the Office of Facilities Planning and Construction (OFPC) of the University of Texas at Austin. The second is the Texas General Service Commission (TX GSC). The third is the Texas Higher Education Coordinating Board (THECB). The three bodies were selected because they are public agencies dealing with building project planning and construction. The three offices are also located in Austin, Texas. In addition, these three offices were extremely helpful and cooperative in providing assistance in information gathering and data collection.

3.4.2.1 Office of Facilities Planning and Construction (OFPC) of the University of Texas at Austin

UT OFPC is responsible for reviewing facilities construction programs before it is forwarded to the chancellor for approval. The OFPC is also in charge of building construction planning and oversees and manages all the construction projects undertaken by all UT components. At the time of data collection, OFPC had just begun the development of the project database, which only extends from 1992 to the present. The information recorded for each project has a high level of detail. The information included the detailed building characteristics, such as gross square footage and number of elevators, as well as the detailed project cost breakdown by various cost categories. However, the database is still considered in its development stage and the total number of new building construction projects in the database is only 32 at the time of this data gathering effort.

3.4.2.2 Texas General Service Commission (TX GSC)

TX GSC is a body that is in charge of planning, acquiring, constructing and operating state buildings. At the time of this research data collection in 2001, the TX GSC had just developed its database for building costs. The database was developed based on the data gathering effort made in 1999, the Commission's Request for Information. The data collection effort by TX GSC is still ongoing, but at the time of data collection for this research in 2001, TX GSC only had 88 projects in its database that was new building construction projects. In addition, the information provided by the database is limited to only the year completed, building use, construction type, square footage, and total costs. Due to the limited information available for each case, the data from TX GSC is deemed unsuitable for this study.

3.4.2.3 Texas Higher Education Coordinating Board (THECB)

THECB was created by the Texas Legislature in 1965 to "provide leadership and coordination for the Texas higher education system to achieve excellence for the college education of Texas students." THECB is a government body overseeing the approval of all public institutions' construction in Texas. Before each institution can build a new building, one of the first steps is to submit a proposal of the project to THECB. The proposal will include the justification of the needs of the project, the funding sources of the project, and the costs associated with the project. UT OFPC data represents a subset of the projects submitted to THECB, and is included in THECB data collected.

From this preliminary investigation, THECB data are the most extensive and have the level of detail that is most suitable for the proposed research study. In addition, the THECB data also encompass only university types of new building construction projects, thus providing a good consistent sample of project data that facilitate this exploratory building parameters and cost investigation. Consequently, THECB data is selected for use in this research analysis and the subsequent cost estimating model development.

As for the cost indexes data, a published cost manual is consulted. The cost manual or cost file is designed to provide construction cost data for project estimating and budgeting purposes. In addition to providing various work items unit cost, the cost file also usually includes the cost indexes. The construction indexes data used in this research are obtained from *Means Building Construction Cost Data 2002*. The primary motivation in using the Means cost file as a data source was the popularity and the availability of the cost file. Most offices use Means cost file, including THECB, UT OFPC, and TX GSC. In addition, the Means cost file also provides City Cost Indexes for many major cities in Texas. Two indexes are extracted and used in this study: Means City Cost Indexes for cities in Texas and Means Historical Cost Indexes. Means indexes are based on a composite building representing current design practices. It has quantity weightings assigned to 66 construction materials, 21 crafts, and 6 types of construction equipment rentals [Means 2001]. The two indexes are used in the normalization of cost data for the projects in the database, which are constructed at various cities in Texas and at different times across the entire period of the data collected.

3.4.3 Data Collection Procedures and Database Development

The historical project data needed for this research is to be extracted from previously submitted THECB Construction Application Forms. However, at the time of the research, THECB had no electronic database for these historical project data. Therefore, one of the first tasks for the researcher is to locate the actual paper forms. Three types of requests are submitted to THECB: request for land acquisition, renovation construction request and new building construction request. Since this research is focused on new building construction, only the Construction Application Forms for new buildings are extracted and used in the development of an electronic database.

After consultation with members of the committees, staffs at UT OFPC and THECB, it was decided that the data gathering effort would extend as far back as 1990. The project applications and construction period for this data collection is thus from 1990 and 2000, an eleven year time span. It is felt that this time period would yield a sufficient number of projects that is relevant and most reflective of the current and future construction projects.

Past project information and cost data were manually collected from 1990 and 2000. The criteria used in selecting projects for inclusion in the database are:

1. New construction project
2. Building construction project
3. Limited to commonly built project and not unique project, such as an observatory construction project

For this research, Microsoft® Access® 2000 is used as the database management software. Three tables were created for recording the past project data set, historical cost index, and city cost index used in this study.

3.5 DATA PREPARATION

Data preparation for this study aims at producing a data file that is to be used for the subsequent data analysis. Data preparation consists of two main tasks: relevant data extraction from the database and normalization of project cost data.

3.5.1 Extraction of Relevant Data

An information database is a collection of tables that are related. Records in two separate tables are connected through a common field that appears in both tables. For this database, the common fields are time (year) and location (city). Generally, the database contains more information than that required for the data analysis. Therefore, the relevant data needed for the research must be extracted.

Relevant data is extracted from the database through a query. A query is like a question that is asked about the data stored in the database. It specifies the data to be extracted. For this research, the project costs need to be normalized, therefore the following information is extracted for each record:

1. Project Identification Number
2. Year Constructed
3. Project Location

4. Historical Cost Index
5. City Cost Index
6. Project Costs

Before the analysis can begin, the data need to be normalized. The query is exported to another software, Microsoft® Excel® 2000, which is a spreadsheet program. The normalization of the cost data is carried out in Excel® to produce normalized cost data.

3.5.2 Normalization of Cost Data

Data normalization is an important process for adjusting the cost data to the same basis. As discussed in the previous section, the data must be normalized for the location differences and the time differences. For this research, all the cost information is normalized to the dollar amount of project constructed in Houston, in the year 2001. The normalization computations are as follows:

Time Adjustment

$$\text{Cost for 2001} = \text{Cost for Year A} \times \frac{\text{Index for 2001}}{\text{Index for Year A}}$$

Location Adjustment

$$\text{Cost in Houston} = \text{Cost in City A} \times \frac{\text{Index for Houston}}{\text{Index for City A}}$$

The normalization process produced the new costs data, whereby all the costs for each project are adjusted to Houston in 2001. This normalized costs data set is then recombined with other project characteristics data so as to produce a data file that will be used in the next stage for data analysis. The project characteristics data extracted from the database and combined with the normalized costs data are as follows:

1. Project Type
2. Gross Square Footage
3. Assignable Square Footage
4. Usage Ratio
5. Number of Floors

This data file is then exported to a statistical analysis program. The Statistical Package for Social Scientists, SPSS® version 10.1 for Windows® is selected for performing the analysis in this research.

3.6 DATA ANALYSIS

Analysis of the data commenced with data examination. Data examination is a time-consuming but necessary step in any statistical analysis method. Careful data examination can lead to better analysis and model construction, which can result in better predictions [Hair et al. 1998]. For this study, the data examination involved determining the characteristics of the sample through various descriptive statistics and using graphical techniques as a means of representing the data. The variables in the data file are examined to assess their frequency distributions, skews, non-normalities, outliers and missing data points. Skews and outliers can create problems even for simple statistics like the mean, therefore their degree and presence must be addressed.

Data transformations will also be used in this research. Data transformations provide a means of modifying variables to correct violations of statistical assumptions or to improve the relationship between the variables [Hair et al. 1998]. Transformation is useful in reducing the skew and pulling in outliers [Hamilton 1992]. The transformations may be based on reasons that are either "theoretical" or "data derived". Theoretical transformations are based on the nature of the data and the logical justifications. Alternatively, data derived transformations are suggested strictly by the data examination. The transformations used in this study are primarily based on the theoretical consideration, so as to facilitate the model explanation and understanding, and to enhance the model acceptance.

After the data have been examined and the transformations have been performed, the dependence technique of multiple regression analysis of the data can be executed. Multiple regression analysis is a statistical technique that can be used to analyze the relationship between a single dependent (criterion) variable and several independent (predictor) variables [Hair et al. 1998]. The objective of multiple regression analysis is to use the independent variables whose values are known to predict the single dependent value. This research uses multiple regression analysis to derive a regression equation or regression model to predict the cost of the building construction in dollar per gross square foot. The independent variables considered are the building types, gross square footage, building usage ratio, and number of floor levels.

As stated in the previous section, SPSS® version 10.1 for Windows® is used to perform all the statistical analysis of the data in this research. The main output for the data analysis is the development of the regression equation or the parametric cost estimating model. This cost model is validated through testing with actual building construction project data. The validation data set is extracted from the main historical project data file and is kept separately for testing and is not used for the model development.

To facilitate the cost model implementation, a computer program is developed to simplify and automate the steps involved in preparing an estimate using the derived model. The next section discusses this computerized approach to the conceptual cost estimating process.

3.7 PARAMETRIC COST ESTIMATING MODEL (PCEM) COMPUTER PROGRAM DEVELOPMENT

The Parametric Cost Estimating Model (PCEM) computer program is built around the validated regression equation derived through the data analysis. The software used for the program development is Microsoft® Visual Basic® Version 6. The program or application developed enhanced the model usability through the graphical user interface (GUI), providing clearly defined input forms, preformatted input lists, automated data retrievals and computations, and standardized cost estimate report generation.

3.8 DOCUMENTATION

The final procedure in the study is the documentation of the research development and findings. This documentation details the research study background, methodology used, data collected and data analysis performed. The development of the computer program is also discussed along with the implementation procedure. The results drawn from the study are also presented, and recommendations for future investigations proposed.

3.9 CHAPTER SUMMARY

The data flow and the research development processes discussed in the previous sections can be summarized in Figure 3.2.

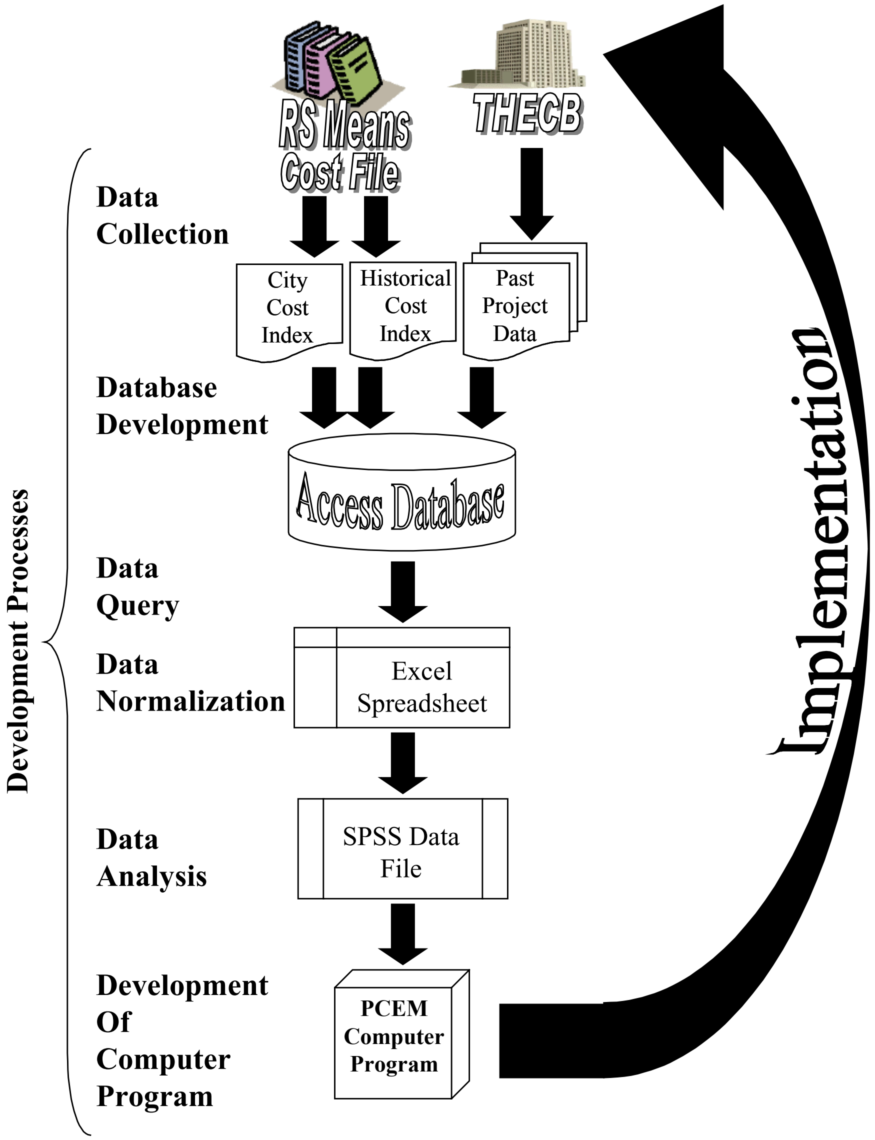


Figure 3.2. Data Flow and Research Development Processes.

This chapter outlined the methodology used in this research study. The data set can be separated into historical project data set and cost indexes data set. The historical project data collection was accomplished by physically visiting THECB and searching through the files to recover the construction project application forms. The cost indexes data were obtained from published *Means Building Construction Cost Data 2002*. An electronic database is developed based on the data collected. The subsequent data preparation involved the extraction of relevant data from the database and the normalizing of the cost data to adjust the costs to a common basis. The base selected is in dollar amount for Houston in 2001. The data analysis for this research involved a comprehensive examination of the data, data transformations, and multiple regression analysis. The output of the multiple regression analysis is the cost estimating model, which forms the heart of the PCEM program. The final procedure for the research study is the documentation of the research effort and results.

CHAPTER 4
DESCRIPTION OF THE DATA

4.1 INTRODUCTION

As discussed in the previous chapters, parametric cost estimating model development is a data intensive process. Data on building characteristics and the associated costs must be collected and statistically analyzed to identify the significant cost drivers and to establish the relationships that form the basis for the parametric cost estimating model development. In addition, cost indexes data must also be collected in order to normalize the past projects data, such that the projects costs are adjusted for the location and time differences in the projects execution.

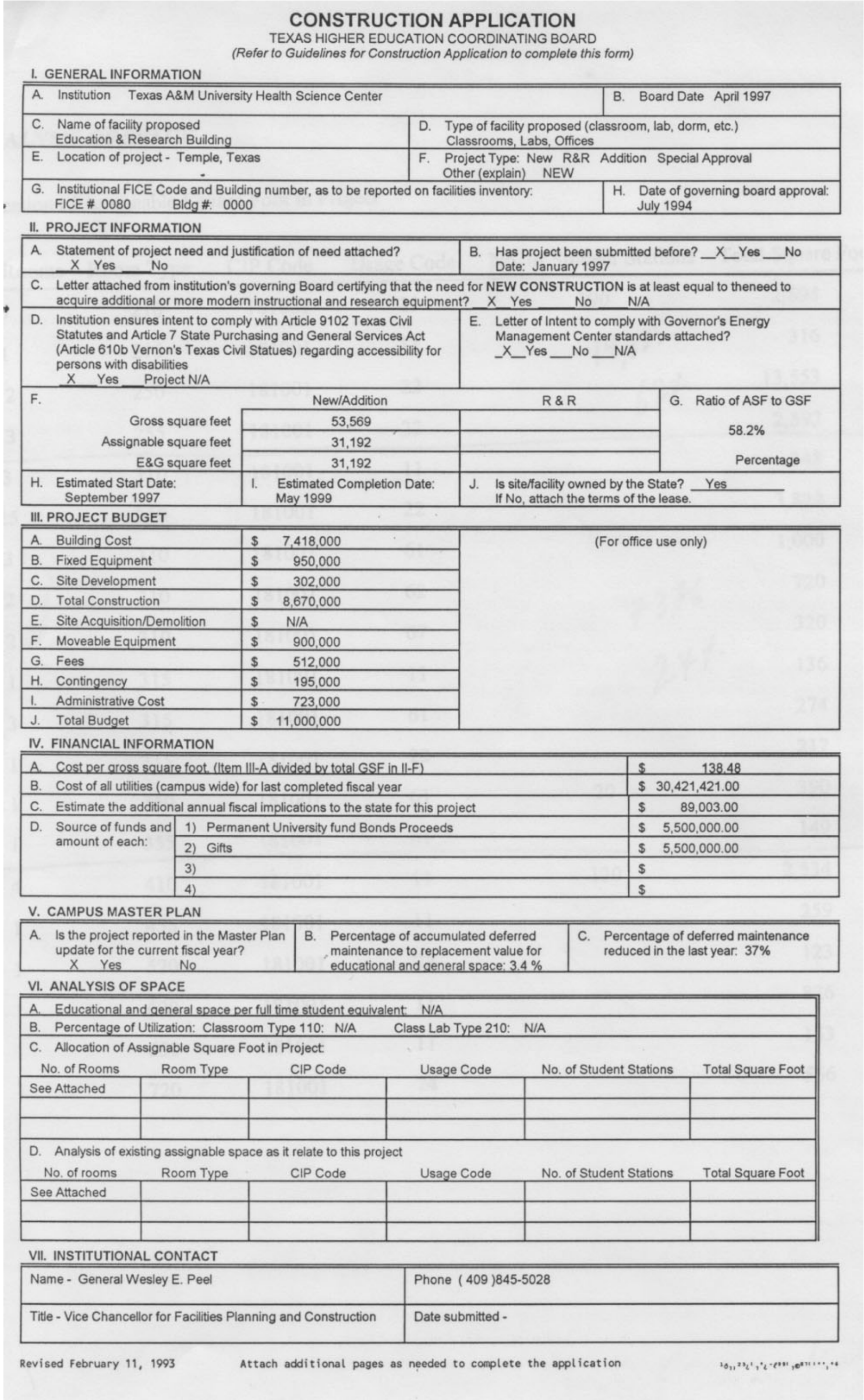
This chapter presents the comprehensive description of the data collected. The data needed for this research can be separated into two parts: historical project data set and cost indexes data sets. Section 4.2 presents the discussion on the historical project data collected, while the cost indexes data sets are presented in Section 4.3.

4.2 HISTORICAL PROJECT DATA SET

The historical project data collected for this research consists of data from 168 new building construction projects constructed at various cities in Texas from 1990 to 2000. The following subsections present the details of the data collected, the general description of the data demographic, and the types of building projects involved.

4.2.1 Description of THECB Construction Application Form

All the project historical data used in this study has been collected from the THECB Construction Application Form and attachments, such as the detailed project description, architectural notes, and architectural drawings. Every public university in Texas seeking to construct a new facility is required to submit the form to THECB for budget approval before the project can be executed. Figure 4.1 shows a sample of a completed THECB Construction Application Form.



CONSTRUCTION APPLICATION

TEXAS HIGHER EDUCATION COORDINATING BOARD

(Refer to Guidelines for Construction Application to complete this form)

I. GENERAL INFORMATION

A. Institution Texas A&M University Health Science Center	B. Board Date April 1997
C. Name of facility proposed Education & Research Building	D. Type of facility proposed (classroom, lab, dorm, etc.) Classrooms, Labs, Offices
E. Location of project - Temple, Texas	F. Project Type: New R&R Addition Special Approval Other (explain) NEW
G. Institutional FICE Code and Building number, as to be reported on facilities inventory: FICE # 0080 Bldg #: 0000	H. Date of governing board approval: July 1994

II. PROJECT INFORMATION

A. Statement of project need and justification of need attached? _X_Yes __No

B. Has project been submitted before? __X__Yes ___No Date: January 1997

C. Letter attached from institution's governing Board certifying that the need for **NEW CONSTRUCTION** is at least equal to theneed to acquire additional or more modern instructional and research equipment? _X_Yes ___No ___N/A

D. Institution ensures intent to comply with Article 9102 Texas Civil Statutes and Article 7 State Purchasing and General Services Act (Article 610b Vernon's Texas Civil Statues) regarding accessibility for persons with disabilities _X__Yes ___Project N/A

E. Letter of Intent to comply with Governor's Energy Management Center standards attached? _X_Yes ___No ___N/A

F.	New/Addition	R & R
Gross square feet	53,569	
Assignable square feet	31,192	
E&G square feet	31,192	

G. Ratio of ASF to GSF 58.2% Percentage

H. Estimated Start Date: September 1997	I. Estimated Completion Date: May 1999	J. Is site/facility owned by the State? __Yes__ If No, attach the terms of the lease.

III. PROJECT BUDGET

A. Building Cost	$ 7,418,000	(For office use only)
B. Fixed Equipment	$ 950,000	
C. Site Development	$ 302,000	
D. Total Construction	$ 8,670,000	
E. Site Acquisition/Demolition	$ N/A	
F. Moveable Equipment	$ 900,000	
G. Fees	$ 512,000	
H. Contingency	$ 195,000	
I. Administrative Cost	$ 723,000	
J. Total Budget	$ 11,000,000	

IV. FINANCIAL INFORMATION

A. Cost per gross square foot. (Item III-A divided by total GSF in II-F)		$ 138.48
B. Cost of all utilities (campus wide) for last completed fiscal year		$ 30,421,421.00
C. Estimate the additional annual fiscal implications to the state for this project		$ 89,003.00
D. Source of funds and amount of each:	1) Permanent University fund Bonds Proceeds	$ 5,500,000.00
	2) Gifts	$ 5,500,000.00
	3)	$
	4)	$

V. CAMPUS MASTER PLAN

A. Is the project reported in the Master Plan update for the current fiscal year? X Yes No	B. Percentage of accumulated deferred maintenance to replacement value for educational and general space: 3.4 %	C. Percentage of deferred maintenance reduced in the last year: 37%

VI. ANALYSIS OF SPACE

A. Educational and general space per full time student equivalent: N/A

B. Percentage of Utilization: Classroom Type 110: N/A Class Lab Type 210: N/A

C. Allocation of Assignable Square Foot in Project:

No. of Rooms	Room Type	CIP Code	Usage Code	No. of Student Stations	Total Square Foot
See Attached					

D. Analysis of existing assignable space as it relate to this project

No. of rooms	Room Type	CIP Code	Usage Code	No. of Student Stations	Total Square Foot
See Attached					

VII. INSTITUTIONAL CONTACT

Name - General Wesley E. Peel	Phone (409)845-5028
Title - Vice Chancellor for Facilities Planning and Construction	Date submitted -

Revised February 11, 1993   Attach additional pages as needed to complete the application

Figure 4.1. THECB Construction Application Form.

The THECB Construction Application Forms are prepared by the institution's office of facilities planning and construction. The forms are prepared at various design stages, from project preliminary design to bid award, but usually before the commencement of construction. Substantial cost changes often involved resubmission of the application. Presently, there is no mechanism for updating the information submitted on the form other than through the application resubmission. Therefore, although the data offers a comprehensive description of the project, problems with their use in the analysis may exist due to the fact that some of the cost information presented may be an estimate of the subsequent cost. Limited information is available about cost overrun and cost changes after initiation of construction. This limitation is further discussed in the limitation of data analysis in Chapter 5.

Information presented by the THECB Construction Application Form can be classified into 7 categories as seen in Figure 4.1. The 7 categories are general information, project information, project budget, financial information, campus master plan, analysis of space, and institution contact. However, for this research, the main focus for data extraction is mainly set on the first three categories. Although, under the general information, the type of facility is sometimes identified, there is no systematic facility classification system. Due to this lack of facility classification standard, the sixth category of information, the analysis of space is consulted. For this research, the classification of facility type is carried out by evaluating not only the listed type, but also by careful examination of the analysis of space and the attached project detailed description.

The general information lists the pertinent project information, while the project information provides the general scope of the project. The project budget presents the list of the various costs associated with the project. Based on *THECB Facility Inventory Procedures Manual*, Appendix F, the details and meanings of the specific items listed under the project information and project budget are described as follows:

4.2.1.1 Project Information

Gross Square Footage Area – is the sum of the floor areas included within the normal outside faces of exterior walls. Integrated in the gross area are the assignable areas, circulation areas, mechanical areas and construction areas.

Assignable Square Footage Area – is part of the gross square footage area which can be assigned or is available for assignment. The assignable areas are associated with the spaces provided for the operation and maintenance of the building. An example of the assignable area as coded by THECB can be seen in Figure 4.4.

Usage Ratio – is the ratio of the assignable square footage and the gross square footage. Usage ratio is a measure of the efficiency of the space usage of the building.

4.2.1.1 Project Budget

Building Cost - includes all costs of construction within five feet of the building line; all items required by codes (fire extinguishers, cabinets, fire alarm systems, etc.); and items normally found in a building regardless of type.

Fixed Equipment - includes all equipment items which may be installed before completion of the building and which are a part of the construction contract, such as lockers, food service equipment, fixed seating, fixed medical equipment, security equipment, state equipment, state lighting, etc.

Site Development - includes all work required which lies within the site boundary and five feet from the edge of the building, i.e., grading and landscape development, athletic fields, walks, site lighting, street furniture, on-site sewage treatment plant, and unusual foundation conditions.

Total Construction - represents the total budget for construction, which is usually the contract documents base bid.

Movable Equipment - This category includes all movable equipment and furniture items, but does not include instructional equipment, i.e., microscopes, library books, etc.

Fees - Costs for architectural and engineering services.

Contingency - A percentage of the total construction cost is included to serve as a planning contingency and construction reserve (change orders, etc.).

Administrative Cost - Items the owner is responsible for during the planning process, i.e., legal fees, site survey, soil testing, insurance, and material testing.

Total Budget - This represents the total budget required to occupy the new facility.

Not explicitly presented in the THECB Construction Application Forms are the numbers of floors for the new buildings. The number of floors for a building is determined by counting the total number of stories above and below grades for the building. For example a 2-story building with a basement is determined to have three floors. Based on literature review the number of floors or building height information has been identified as one of the important parameters used in a parametric cost estimating model. Therefore, the researcher attempts to collect the number of floors data through careful review of the project description and any other information that may be included as part of the construction application, such as architectural drawings. In cases where the information is not available, the individual institution's office of facilities planning and construction are contacted, by both phone calls and e-mails. Through this effort, most of the building's number of floors had been identified, with few cases where the researcher was unable to get any information.

This subsection provides a comprehensive discussion on the details of the THECB Construction Application Form. In addition, the relevant data in the form are identified and the associated descriptions and meanings presented. With this comprehensive layout of the data involved, the following subsection presents the project demographic information and project types involved.

4.2.2 Project Demographics

The historical project data consists of 168 new building construction projects. The cumulative total gross square footage of the 168 projects is 15,740,417 square feet. The cumulative total assignable square footage is 9,620,249 square feet. The cumulative total project value is over $ 2.5 billion (2001 dollars).

The projects are submitted by eight systems of higher education. The eight systems are:

1. The University of Texas System (UT)
2. Texas A&M System (TX A&M)
3. The Texas State Technical College System (TSTC)
4. The State University System (State)
5. The Texas Tech University System (TX TECH)
6. The University of Houston System (UH)
7. The University of North Texas System (UNT)
8. Independent academic unit (Independent)

Figure 4.2 shows the number of projects constructed by the various university systems. The two major systems with the most projects are the University of Texas System and the Texas A&M System. About 35% and 26% of the projects are submitted by the University of Texas System and the Texas A&M System respectively.

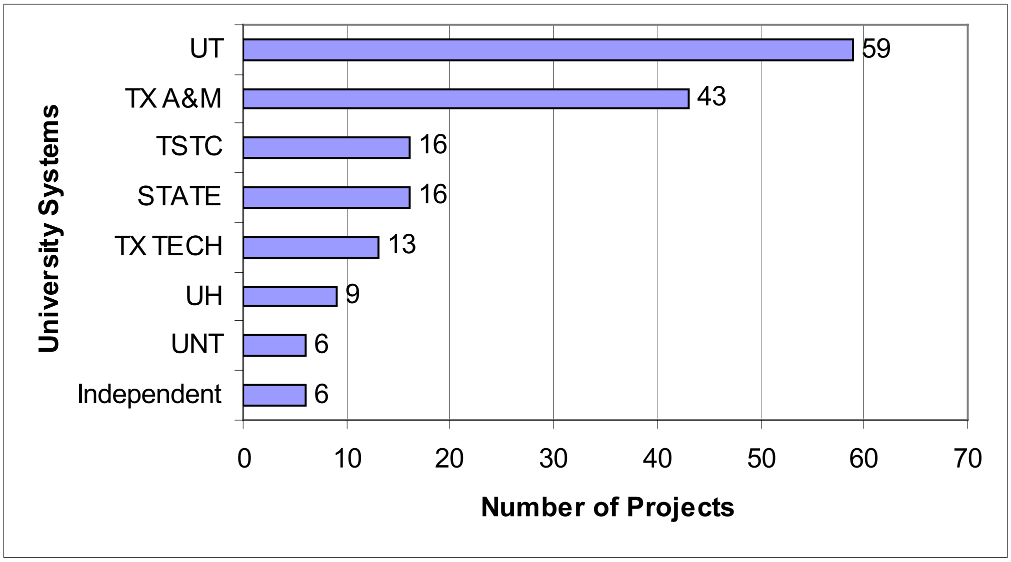


Figure 4.2. Number of Projects Constructed by Various University System.

The breakdown of the submitted projects by various academic units can also be seen in Table 4.1.

Table 4.1. List of Academic Units and Number of Projects

Instituition	Count
Angelo State University	2
Lamar University - Beaumont	1
Lamar University - Orange	1
Lamar University - Port Arthur	2
Sam Houston State University	2
Southwest Texas State University	7
Sul Ross State University	1
Tarleton State University	3
Texas A&M College Station	13
Texas A&M Commerce	1
Texas A&M Corpus Christi	5
Texas A&M Galveston	1
Texas A&M Health Science Center	2
Texas A&M International	5
Texas A&M Kingsville	2
Texas A&M Prairie View	3
Texas A&M Texarkana	1
Texas A&M West Texas	1
Texas Agricultural Experiment Station	3
Texas Engineering Experiment Station	3
Texas Southern University	5
Texas Tech Health Science Center	4
Texas Tech University	9
Texas Woman's University	1
TSTC Amarillo	2
TSTC Harlingen	8
TSTC Sweetwater	2
TSTC Waco	4
UH Downtown	2
UH Fort Bend	1
UH Main	6
UNT	4
UNT Health Science Center	2
UT Arlington	2
UT Austin	10
UT Brownsville	1
UT Dallas	8
UT El Paso	3
UT Health Science Center	6
UT MD Anderson	2
UT Medical Branch Galveston	1
UT Pan American	8
UT Permian Basin	2
UT San Antonio	10
UT Southwestern Medical Center	4
UT Tyler	2

The distribution of the projects across the eleven-year period can be seen in Figure 4.3. Most of the construction activities are concentrated in the last two years, 1999 and 2000, which account for more than 40% of the total submitted projects.

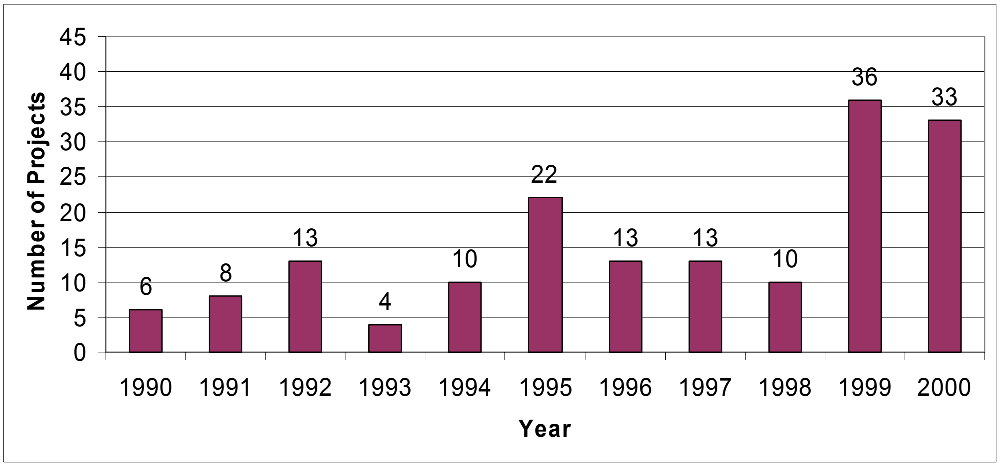


Figure 4.3. Distribution of the Projects Constructed During the 1990-2000 Period.

Table 4.2 lists the 42 cities where the projects are located. It is evident that the cities with high concentration of projects are either major Texas cities, such as Houston and Dallas, or major college cities, such as College Station and Lubbock. Houston, Dallas and San Antonio are the cities with the some of the highest numbers of projects, which are 16, 10, and 13, respectively. While the college cities, which are College Station and Lubbock also have significant numbers of projects, which are 14 and 11, respectively.

Table 4.2. Location of Projects at Various Cities in Texas.

City	Count
Alpine	1
Amarillo	3
Arlington	2
Austin	10
Beaumont	1
Brownsville	2
Canyon	1
College Station	14
Commerce	1
Corpus Christi	5
Dallas	10
Denton	5
Edinburg	8
El Paso	3
Fort Bend	1
Fort Stockton	1
Fort Worth	2
Galveston	2
Harlingen	9
Houston	16
Huntsville	2
Kingsville	2
Laredo	6
Longview	1
Lubbock	11
Midland	1
Odessa	2
Orange	1
Overton	1
Port Aransas	1
Port Arthur	2
Prairie View	3
Richardson	2
San Angelo	2
San Antonio	13
San Marcos	7
Stephenville	3
Sweetwater	2
Temple	3
Texarkana	1
Tyler	1
Waco	4

4.2.3 Type of Facility

As stated in Section 4.2.1, THECB currently has no set standard in classifying the type of facility to be constructed. Therefore, to standardize the type of facility involved for the project cases in the data set used in this research, additional information are investigated. The first and most important consideration is the type of facility listed on the application form. The second source of information is the sixth category of information listed in the application form, the analysis of space. The analysis of space provides the breakdown of the assignable square feet by room type and classification codes. The codes can be found in the *THECB Facility Inventory Procedures Manual*, Appendix C. Some examples of the code used are illustrated in Figure 4.4.

Classroom Facilities		
	110	Classroom
	115	Classroom Service
Laboratory Facilities		
	210	Class Laboratory
	215	Class Laboratory Service
	220	Special Class Laboratory
	225	Special Class Laboratory Service
	230	Individual Study Laboratory
	235	Individual Study Laboratory Service
	250	Non-Class Laboratory
	255	Non-Class Laboratory Service
Office Facilities		
	310	Office
	315	Office Service
	350	Conference Room
	355	Conference Room Service

Figure 4.4. Sample THECB Standard Room Type Codes.

The third source of information considered is the detailed description of the project, which is often submitted along with the application form. The detailed description document generally provides more information about the project by describing in the detail the scope of the project and the intended use of the facility. Considering these three sources of information, the facility type is classified into the following 14 categories:

1. Administrative/Office Building
2. Apartment Building
3. Athletic Facility
4. Classroom Building
5. Dormitory
6. Healthcare Facility
7. Laboratory Storage Facility
8. Library
9. Multi-Purpose Facility
10. Parking Garage
11. Performance Facility
12. Research Laboratory
13. School
14. Student Center

The number of projects classified by these facility types and the facility type composition of the data set can be seen in Figures 4.5 and 4.6, respectively.

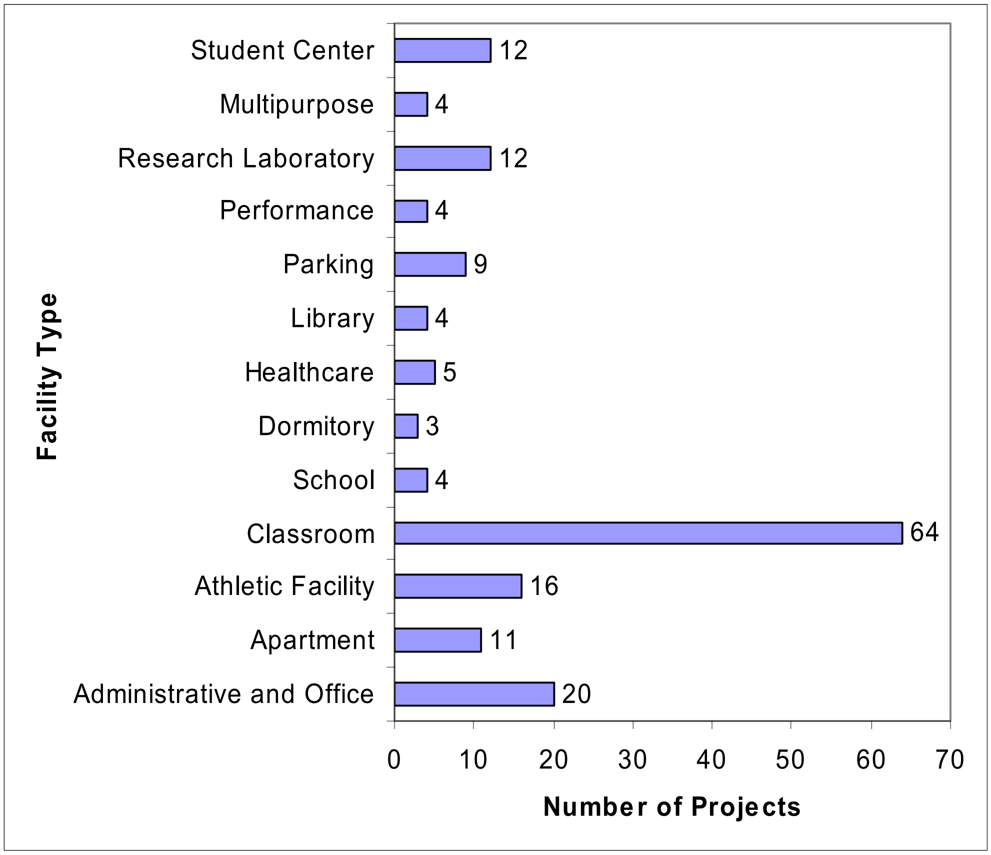


Figure 4.5. Number of Projects by Facility Type.

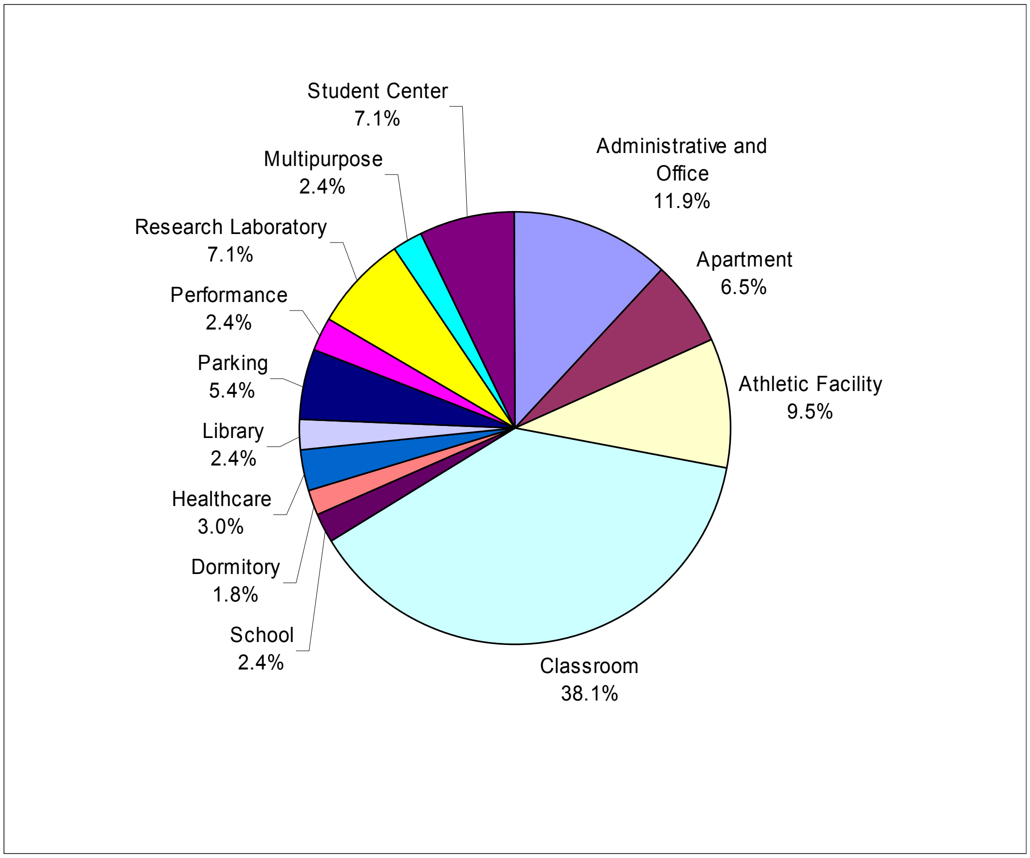


Figure 4.6. Facility Type Composition of Project Data Collected.

It is evident that the major portion of the facility type is classroom, comprising about 38% of the project data collected.

Figure 4.7 illustrates the distribution of the project by the number of floor levels. Most of the facilities are one to three floors, and that less than 5% of the projects have more than five floors.

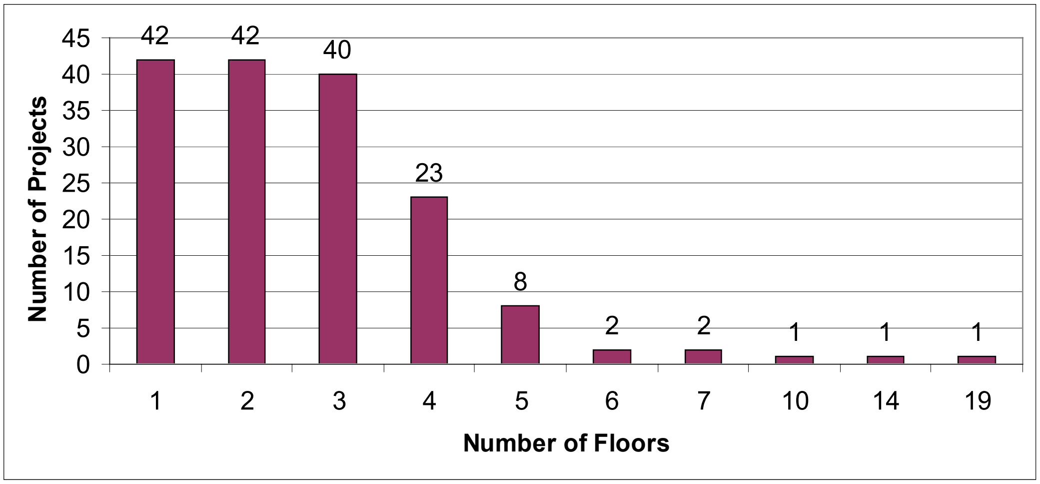


Figure 4.7. Project Distribution of the Number of Floor Levels.

4.3 COST INDEXES DATA SET

The construction indexes data used in this research are obtained from *Means Building Construction Cost Data 2002*. The two indexes extracted are Means Historical Cost Indexes and Means City Cost Indexes for cities in Texas. The original historical cost indexes obtained are based in 1993 and are subsequently adjusted to 2001. The historical indexes can be seen in Table 4.3.

Table 4.3. Historical Cost Indexes.

Historical Cost Indexes		
Year	Index Jan 1, 1993 = 100	Index July 1, 2001 = 100
July 2001	125.1	100.0
2000	120.9	96.6
1999	117.6	94.0
1998	115.1	92.0
1997	112.8	90.2
1996	110.2	88.1
1995	107.6	86.0
1994	104.4	83.5
1993	101.7	81.3
1992	99.4	79.5
1991	96.8	77.4
1990	94.3	75.4

Adapted from [Means 2001]

The city cost indexes as listed in Means are based on the average of 30 major U.S. cities. The base for the city cost indexes are adjusted to Houston, which has the highest index among the Texas cities. Unfortunately, Means does not have the indexes for all the Texas cities encountered in this study. Consequently, the index for the nearest neighboring city is used as the substitute index for a city with a missing cost index. The city cost indexes can be seen in Table 4.4.

Table 4.4. City Cost Indexes.

City Cost Indexes		
City	**Index National Average = 100**	**Index Houston = 100**
Alpine*	76.3	86.4
Amarillo	80.8	91.5
Arlington*	82.4	93.3
Austin	81.3	92.1
Beaumont	83.4	94.5
Brownsville*	77.3	87.5
Canyon*	80.8	91.5
College Station	82.4	93.3
Commerce*	78.3	88.7
Corpus Christi	78.8	89.2
Dallas	85.2	96.5
Denton	78.3	88.7
Edinburg*	77.3	87.5
El Paso	78.1	88.4
Fort Bend*	88.3	100.0
Fort Stockton*	76.3	86.4
Fort Worth	82.4	93.3
Galveston	86.5	98.0
Harlingen*	77.3	87.5
Houston	88.3	100.0
Huntsville	73.1	82.8
Kingsville*	78.8	89.2
Laredo	77.3	87.5
Longview	73.7	83.5
Lubbock	79.9	90.5
Midland	78.6	89.0
Odessa	76.3	86.4
Orange*	83.4	94.5
Overton*	79.9	90.5
Port Aransas*	78.8	89.2
Port Arthur*	83.4	94.5
Prairie View*	82.4	93.3
Richardson*	85.2	96.5
San Angelo	75.5	85.5
San Antonio	82.6	93.5
San Marcos*	81.3	92.1
Stephenville*	79.8	90.4
Sweetwater*	78.8	89.2
Temple	76.2	86.3
Texarkana	76.7	86.9
Tyler	79.9	90.5
Waco	79.8	90.4

Adapted from [Means 2001] (* Index based on nearest city)

4.4 CHAPTER SUMMARY

This chapter presents the general descriptions of the data collected for this research. The two data sets collected are the project historical data set and the cost indexes data set. The details of the THECB Construction Application Form are presented to provide the background information for the subsequent discussions on the historical project data collected. The project data are a collection of new building construction projects executed between 1990 and 2000. The projects are located across 42 cites in Texas, with greater concentration in the big cities and college cities. Most of the projects are submitted by the two major systems, the University of Texas System and the Texas A&M System. Thirty eight percent of the data collected are classroom type facility. Most of the buildings have one to three floors, and less than 5% of the projects have more than 5 floors.

The cost indexes used in this study are the Means Historical Cost Indexes and Means City Cost Indexes for cities in Texas. The two indexes are adjusted to a new base, which is for year 2001 and Houston, respectively.

The next chapter focuses on the data analysis for this research.

CHAPTER 5
DATA ANALYSIS AND PRESENTATIONS OF RESULTS

5.1 INTRODUCTION

A parametric method in conceptual cost estimating involves the identification of the significant building parameters, or cost drivers, and the development of the parametric model or equation. The parametric model is a mathematical model or function, which defines the cost estimating relationships (CERs). The word parameter, as used in the CERs context, refers to building characteristics, such as floor area and building height. It has a different meaning than the commonly known term used in statistics, which refers to the coefficients found in a regression equation.

A parametric model is typically a statistical model that characterizes the cost of a project or task as a function of one or more independent variables, which are usually the scope or design related parameters. A parametric model can be used to infer the significance and relative importance of building parameters used in the model. Since the model is developed from relevant past project data, the success of the model depends on the ability of the model to capture the past pattern in the data. In this manner, the parametric model developed can also be used to estimate the cost of a future project.

Chapter 4 presents an introduction of the two data sets collected. The cost indexes data sets are used in the normalizing of the project costs data, thereby producing the normalized data file that will be used in the analysis stage. This chapter focuses on this analysis stage, specifically the data examinations and analyses of the building parameters data and the normalized costs data. The analysis tools and statistical techniques used are first presented. The variables selection issue is then discussed, follow by a discussion on the definition of the population to be studied. The details of the data examination and variable transformation are also presented before the comprehensive discussion on the multiple regression analysis and the parametric cost estimating model development. In closing, the limitations of the analyses are also discussed.

5.2 ANALYSIS TOOL AND STATISTICAL TECHNIQUES

Statistical analysis permits researchers to reach tentative conclusions about the existence and strength of any relationships of concern [Knoke and Bohrnstedt 1994]. The purpose of the data collection is to produce statistics. A statistic is some characteristic of the sample [Norušis 1995]. For this study, all statistical analyses were performed utilizing the statistical package, SPSS® for Windows® Version 10. The historical project data collected for this study is carefully examined and statistically analyzed with various graphical techniques and descriptive statistics. The subsequent development of the parametric cost estimating model is performed using the statistical technique of multiple regression analysis. Understanding the various analysis techniques and being able to interpret the results are imperative to the success of the research. As such, the following sections describe several statistical techniques employed in this research study.

5.2.1 Graphical Techniques for Data Examination

The following graphical techniques are used extensively in this study to explore the nature of the data collected.

5.2.1.1 Histogram

The starting point for data examination is usually the histogram. A histogram is a graphical representation of a single variable that denotes the frequency of occurrence within data categories [Hair et al. 1998]. A histogram is used to characterize the shape of the distribution for a single variable.

5.2.1.2 Scatterplot

A scatterplot is a two-dimensional graph on which the joint values for two metric variables are plotted. In this way, a scatterplot is a representation of the data based on two variables. It is used to visually check, detect, or confirm the relationships or any patterns of association between two variables.

5.2.1.3 Boxplot

A graphical technique called the boxplot is extremely useful in exploring the characteristics of the data. Figure 4.1 provides an annotated sketch of a boxplot. A boxplot is a display that summarizes information about the distribution of values. Specifically, boxplots are used in this study to reveal the summary statistics for the variable distributions associated with each facility type.

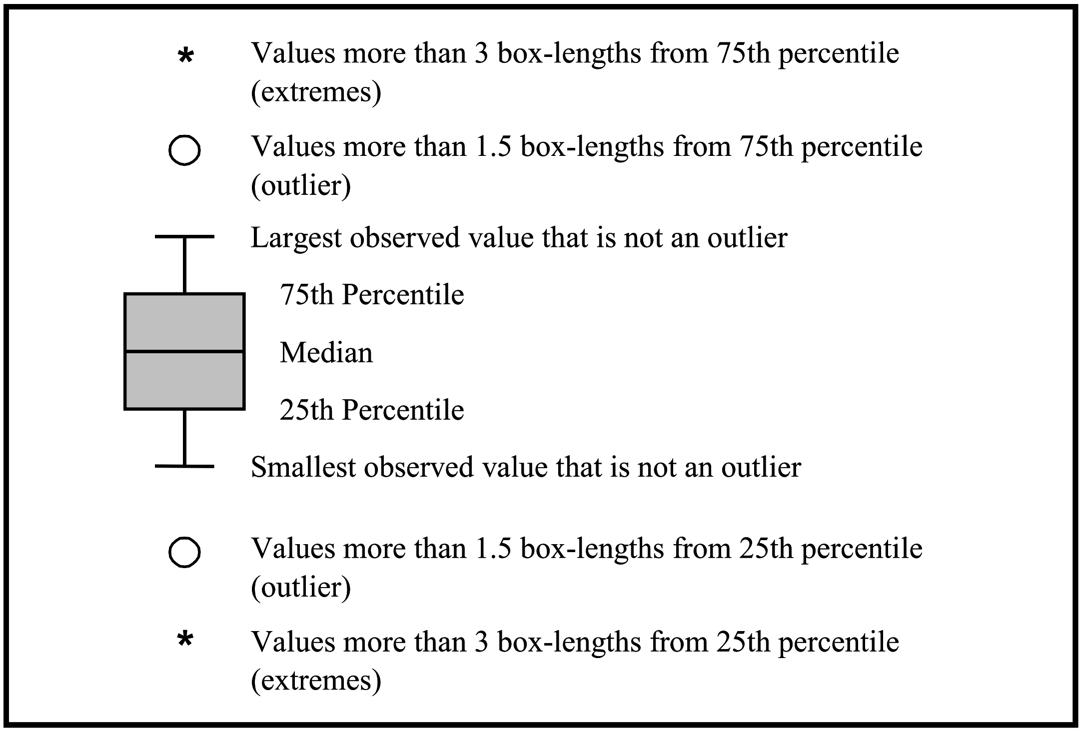


Figure 5.1. Annotated Sketch of the Boxplot.

5.2.2 Multiple Regression Analysis

Regression is referred to as the “core technique” of modern statistical analysis [Hamilton 1992]. Regression analysis is by far the most widely used and versatile dependence technique, applicable in every facet of business decision-making. It is most useful in relating a factor (or factors) to a specific outcome. It is the foundation for most business forecasting models, and is widely applied to most decision making tools ranging from marketing and feasibility study, to program evaluation [Hair et al. 1998]. These widespread uses and acceptance demonstrate that regression analysis is a powerful analytical tool design to explore all types of dependence relationships.

Multiple regression analysis is a statistical technique used to analyze the relationship between a single dependent variable and several independent variables. The basic formulation is:

$$\hat{Y} = b_0 + b_1 V_1 + \ldots\ldots + b_n V_n$$

where $\hat{Y}$ = Predicted dependent variable

b_0 = Constant

b_1 = Partial Regression Coefficient of V_1

V_1 = Independent Variable 1

b_n = Partial Regression Coefficient of V_n

V_n = Independent Variable n

n = number of variables

The partial regression coefficient (b) measures the amount of change in the dependent variable for a one-unit change in the independent variable. All the partial regression coefficient values will be determined through the method of ordinary least squares (OLS). That is, the selections of the coefficients are such that the resulting model produces the smallest sum of squared differences between the observed and predicted values of the dependent variable. Any other coefficients used will yield a larger sum of squared residuals.

The measure for the goodness-of-fit for the model is the coefficient of determination, or R^2, which represents the proportion of dependent variable variance that is accounted for by its linear relationships with the independent variables. To determine whether or not the model with its independent variables is a significant predictor of the dependent variable (i.e., significant test for R^2), a statistical significance test is performed using the F-statistic. The R^2 value is statistically significant, or different from zero, if the computed F-statistic is greater than the F-critical value for the defined probability level. For this study, the p-value used is at a 0.05 level. Therefore, for the output interpretation, if the obtained significance level (p-value) associated with the F-statistic is less than 0.05 (at 95 percent confidence), then R^2 is statistically significant.

5.3 VARIABLES SELECTION FOR MODEL DEVELOPMENT

The first step to multiple regression analysis is the selection of the variables to be used in the analysis. Multiple regression analysis is a dependence technique, therefore the variables to be used as the dependent variable and the independent variables must be specified. The selection of both types of variables should be based principally on conceptual or theoretical grounds [Hair et al. 1998]. As in all functional estimating models, there must be logical relationships of the independent variables to the dependent variable.

5.3.1 Variables Suggested in Previous Study

In-depth discussion on related previous research in parametric cost estimating can be seen in Chapter 2. All the methods and models identified acknowledge the time and location influences on project cost. Kouskoulas and Koehn have added the location and time variables to their model. Bowlby and Schriver also added the location variable into their model, while using cost indexes to normalize the cost. Karshenas, on the other hand, chooses to normalize the project cost for both the time and location, to New York City, March 1982.

In this study, the data are derived from projects located in 42 Texas cities, across the 11 year time period, 1990-2000. However, their distributions are not evenly distributed, with most projects concentrated in major cities or college cities. In addition, a major portion of the projects is also derived from projects constructed during the last two year of the data collection period, in 1999 and 2000. Due to this unevenness in distribution, and the very limited representations of the data for most cities and year, the time and location influences on the costs data are accounted and adjusted for through the data normalization process as discussed in Chapter 3, and are not explicitly identified in the model.

Based on the literature review, a parameter associated with the height of the building or the number of floors is identified as a significant parameter to be used in the parametric cost estimating model [Bowlby et al. 1986, Karshenas 1984, Kouskoulas et al. 1974]. In addition, parameters associated with building floor areas are also identified as significant [Bowlby et al. 1986, Karshenas 1984]. In developing the cost estimating model, Karshenas uses the typical floor area, while Bowlby and Schriver use the total floor area of the building. These height and size related parameters provide some measures about the scale of the building.

End use of the building, or building type is also identified as the significant parameters to be considered in the cost estimating model development [Bowlby et al. 1986, Kouskoulas et al. 1974]. However, Kouskoulas also states that “the fact remains that if the methodology is applied to classes of buildings instead of to the whole population of buildings, one is bound to get very good results.” Therefore, in order to produce a successful cost estimating model, this issue about the building type and the target population of building to be considered must be taken into account in the model development.

Although the building type may distantly reflect the quality measure of the construction at a macro level, applying to model concerns with multiple building types or heterogeneous population, there is a lack of a quantitative quality measure for a more homogeneous population of building type construction. To address the quality aspect of the building project, Kouskoulas and Koehn developed a quality index based on the 1) quality of workmanship and materials used in the construction process, 2) building use, 3) design effort, and 4) material type and quality used in various building components. However, the application of this indexing can be a subjective process and demand information not usually recorded from past data or cannot be readily determined at the conceptual level for a new project. Alternatively, Bowlby and Schriver use both the building type and structural framing type to account for the quality of the construction.

5.3.2 Variables Selection Considerations

For this study, the selection of the parameters to be used in the model development is based on the available project historical data and the significant parameters suggested by the literature review. In addition, since the intent of the model is not only for the identification and explanation of the parameters affecting the construction cost, but also for use in estimating the construction cost, the following variable characteristics are also required in variable selection:

1. Variable should be well established with a clear definition so as to minimize ambiguity and inconsistency;
2. Variable should have readily quantifiable or measurable values;
3. Variable values should be available with reasonable accuracy in the early project stages when conceptual cost estimating is required.

With the above considerations on desirable variable characteristics and suggestions from literature reviews, the next section discusses the variables selected for this study.

5.3.3 Variables Selected for this Study

Considering the above discussions on previous research and desirable variable characteristics, the historical data collected is examined and the variables for this study are selected as follows:

1. Building gross square footage (GSF)
2. Number of floor levels (FLOOR)
3. Space usage ratio (RATIO)

4. Building cost per gross square foot of floor area ($perGSF)

These building parameters selected are quantitative measures of the building characteristics that are well established with clear definition. These parameters are also usually the first few parameters identified for the building construction project. Building gross square footage and number of floor levels are measures of the scale of the project. These variables have been shown to significantly affect the building construction cost, and are thus considered in this study. In addition, this study proposes a new variable, the space usage ratio, to be investigated for potential application in cost estimating model development. The space usage ratio is the ratio of the assignable square footage area and the gross square footage area. It is a crude measure of how the building floor spaces are utilized. More importantly, it can also be used to infer the efficiency of the building design, building usage type, and ultimately the building construction quality. For example, low usage ratios are usually related with more expensive high occupancy facilities whereby more spaces are required for circulation areas and mechanical areas. These high occupancy facilities also tend to cost more due to the need for better design, more durable construction, and higher demand for HVAC, plumbing, and electrical requirements. It is anticipated that this new measure of building quality will have a significant correlation with the building cost, and can subsequently be used as a parameter for estimating the building cost. In addition, this new measure of building quality will also be applicable across various facility types.

The unit measure of cost or cost per square foot, specifically building cost per gross square foot of floor area ($perGSF), is selected as the dependent variable. This unit measure is determined by dividing the building cost by the gross square footage area of the building. This choice is supported by literature reviews, judgment on a good cost measure of buildings, and data availability. As mentioned previously in Chapter 2, the most common unit cost estimate for buildings is the cost per square foot. The unit method of cost per square foot is also the most popular of all estimating methods. The building cost is selected instead of the total project cost due to the need to isolate the construction cost to just the building construction cost and minimize any variations that may be location or project specific, such as the site development cost, contingencies, and administrative costs. It is anticipated that by using just the building cost, the variation in the cost can be better explained by the selected building parameters.

5.4 DEFINING THE TARGET POPULATION AND DATA SET PREPARATION

As stated in Section 5.3.1, the building type can be an influential factor affecting the unit cost of the project. That is, the cost per square foot for two buildings of the same floor area and floor levels but of different types can be considerably different. In this manner, the issues about the building type in the population must be addressed.

One of the first steps in designing a research project is to define the population to be studied. The entire population of the project types would be too large and diverse to study and would produce questionable results because of this diversity. Conclusions reached by a study of such a wide mix of project types would provide overly general information that would not prove useful to any specific subset of the population of project type. As a result, the sample was narrowed to project types that are not drastically different with roughly similar unit cost.

5.4.1 Data Screening with Boxplots

Data screening is performed on the data set in order to examine the distributions of the variables selected based on the different building types. The purpose of data screening is to identify any significant deviations in the distributions of the variables based on the building types.

Data screening with respect to various building types are performed using the boxplots. Four boxplots are made based on the building's GSF, FLOOR, RATIO, and \$perGSF. The four boxplots can be seen in Figure 5.2, 5.3, 5.4, and 5.5.

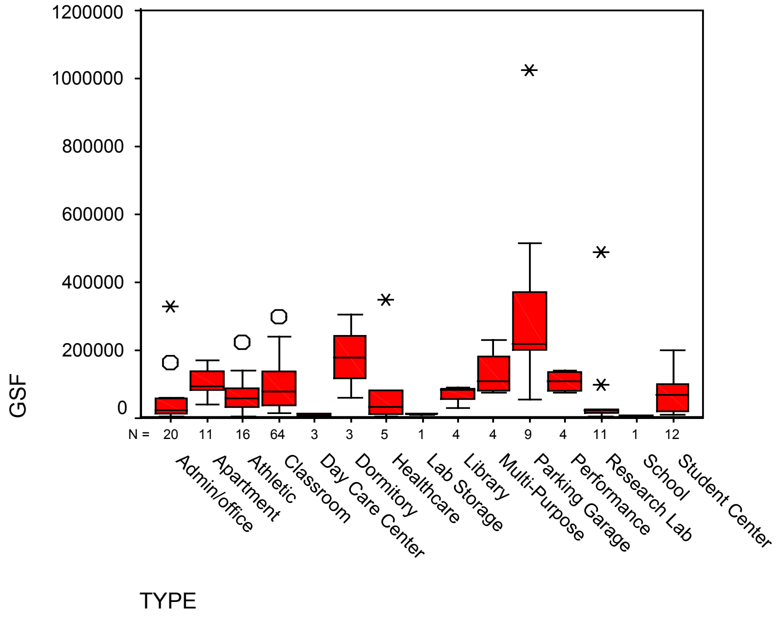


Figure 5.2. Comparison of Gross Square Footage by Building Type (N=168).

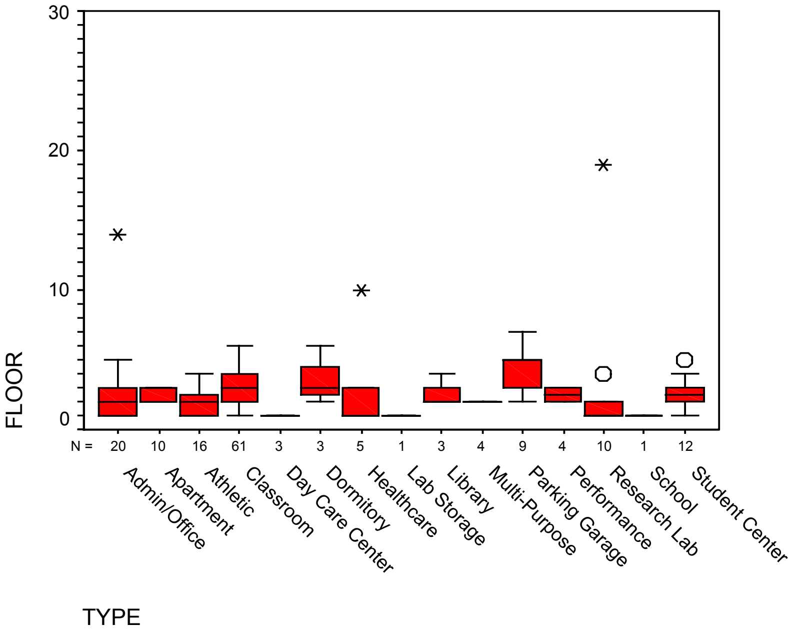


Figure 5.3. Comparison of Number of Floor by Building Type (N=162).

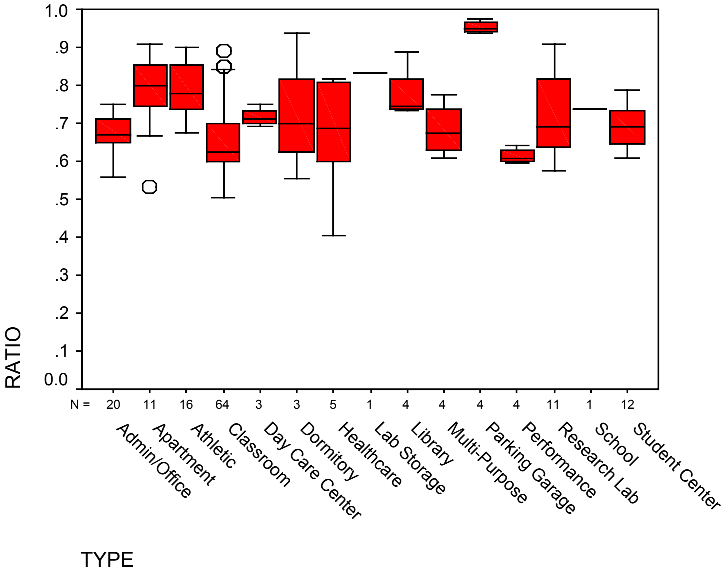


Figure 5.4. Comparison of Usage Ratio by Building Type (N=168).

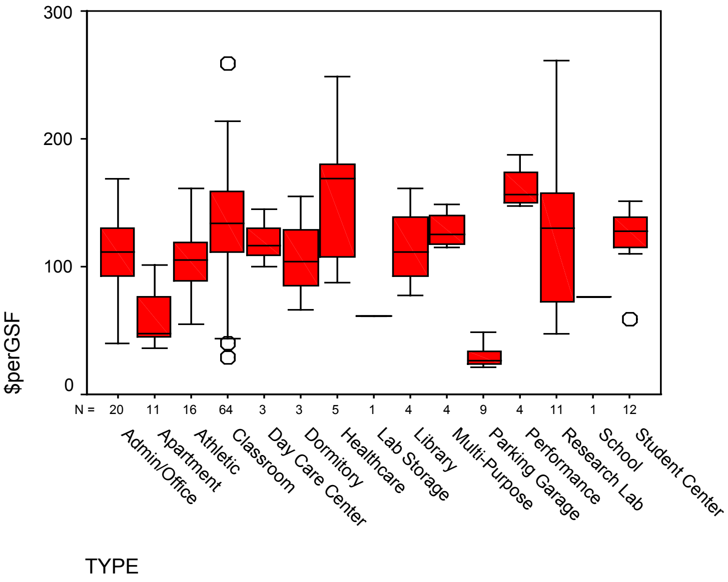


Figure 5.5. Comparison of Building Cost per GSF by Building Type (N=168).

As previously discussed and shown in the boxplots above, a major portion of the project data is classroom type facilities. It can also be seen that the parking garage is significantly different from other facilities. The gross square footage area for the parking facility is generally more than other facilities. The usage ratio for the parking garage is also significantly higher than the other facility types. More importantly, the building cost per gross square footage is significantly lower than the other facility types considered. This is logical. A parking garage is designed to contain automobiles, thus it is usually a large facility. It is a specialized facility with a single use. It is also a simple facility with minimal finishing work, requiring very little mechanical and electrical installations. In this way, parking garages tends to have a high usage ratio and generally a lower unit cost. Consequently, due to these unique characteristics of the parking garage projects, the decision was made to remove the projects from the data set so as to produce a more homogeneous project population.

Figure 5.5 shows that the apartment type facility also has a significantly lower unit cost than the other facility types. Considering the fact that an apartment type project is different from other institution type projects, such as classroom, the apartment type facility will also be similarly removed from the data set. In addition, upon closer examination, several storage or warehouse related facility types are also identified. Due to the simpler nature of construction involved, these projects tend to have low unit costs. Therefore, they are also removed from the data set. Table 5.1 lists all the deleted projects discussed.

Table 5.1. Deleted Projects.

Project ID	Project Name	Facility Type
18	Parking Garage No. 2	Parking Garage
36	Parking Garage at West 7th & Lavaca	Parking Garage
63	Parking Garage at Student Center	Parking Garage
70	Parking Garage No.3	Parking Garage
102	Commuter Parking Structure	Parking Garage
117	Parking Garage No. 4	Parking Garage
149	Parking Garage	Parking Garage
160	West Campus Parking Garage	Parking Garage
180	Parking Garage South	Parking Garage
27	Student Housing	Apartment
53	Student Housing	Apartment
81	Student Apartment	Apartment
87	Residence Facilities-16 buildings	Apartment
104	Waterview Park Apartments-Phase VI	Apartment
135	Waterview Park Apartments Phase VII	Apartment
165	Student Housing	Apartment
167	Student Housing-Phase VIII	Apartment
168	Student Housing	Apartment
172	Student Housing	Apartment
173	Student Housing	Apartment
25	Autobody/Automotive Technology Building	Classroom
16	Field Lab, Animal Facility and Storage	Laboratory Storage
17	Balcones Research Center-Warehouse Building	Library
19	Texas Beef Industry Center	Research Laboratory
3	Research Laboratory-Support Building	Research Laboratory

In summary, 25 projects are removed from the data set, thereby reducing the data set from 168 projects to 143 projects.

5.4.2 Missing Data

From the truncated data set (N=143), further data examination reveals four projects with missing data on the number of floors. The four projects are listed in Table 5.2.

Table 5.2. List of Projects with Missing Data on Number of Floor.

Project ID	Project Name	Facility Type
89	C.A.Bassett Lab-Pulse Lab Addition	Research Laboratory
112	Athletic Academic Services Building	Classroom
134	Athletic Training/Rehabilitation Center	Classroom
182	Experimental Science Building	Classroom

Efforts have been made to contact the respective institutions, both by telephone and e-mails, but to no avail. Due to the time constraint of this research, the numbers of floors for these projects are classified as missing. The simplest and most direct approach for dealing with missing data is to include only those observations with complete data, also known as the complete case approach [Hair et al. 1998]. The complete case approach is suitable for instances in which the extent of the missing data is small and that the sample is large enough to allow for these deletions. Consequently, these four projects with missing data are removed from the data set, reducing the data set to 139 projects.

5.4.3 Description of the New Project Data Set

The resulting project data set can be described by building types as seen in Table 5.3. Due to the limited sample size of the data collection, and the limited representations of the projects in some building types, the data set will no longer be sub-divided. Consequently, the analysis will be performed on this entire data set and generalized for these various types of university buildings.

Table 5.3. List of Building Types in the New Data Set (N=139).

Building Type	# of Project
Administrative and Office	20
Athletic Facility	16
Classroom	60
School	4
Dormitory	3
Healthcare	5
Library	3
Performance	4
Research Laboratory	8
Multipurpose	4
Student Center	12

5.4.4 Extraction of Data for Validation

With the data file refined and ready for analysis, the next step is to extract the projects to be used for future model validation. Seven validation projects, representing about 5% of the population, are randomly selected and extracted from the data set. These validation projects are not used in any analysis steps, and are reserved for the subsequent model validation step. The seven projects extracted are listed in Table 5.4.

Table 5.4. Validation Projects.

Project ID	Project Name	Facility Type
50	University Services Center	Administrative/Office
43	Special Events Center	Multi-Purpose
2	Health Science Building	Classroom
23	Allied Health Technology Building	Classroom
48	Early Childhood Center	Classroom
72	Student Health Center	Healthcare
26	Student Development Center	Student Center

The final data set for analysis therefore contain 132 projects.

5.5 DATA EXAMINATION AND MANIPULATIONS

This section focuses on the data examinations and data manipulations performed on the data set. Data examination is an essential part of any multivariate analysis technique. It allows the researcher to attain a basic understanding of the data and relationships between the variables. Data manipulation, or transformation, provides a means to modify variables. The modification may be required for one of two reasons: (1) to correct violations of the statistical assumptions, or (2) to improve the relationship between variables [Hair et al. 1998]. The discussions on data examination are first presented, which is followed by the discussion on the data transformation used in the analysis.

5.5.1 Univariate Analysis of the Variables Selected

Data examination for this study begins with the univariate analysis, or the study of variables one at a time. Table 5.5 illustrates the descriptive statistics computations for the four variables considered in the study. From the table, it can be seen that the FLOOR ranges from one floor to 19 floors. However, the mean for the floor level is at 2.74 and the skewness is 4.226, indicating that the number of floors for the data set has a positive skew (refer to Figure 5.6). The GSF ranges from 3,360 SF to 485,740 SF, and the mean is 82,310 SF. The GSF also have a slight positive skew of 1.982 (refer to Figure 5.7). The skews in the FLOOR and GSF data are expected as generally there are more small buildings, with one to three floors, than there are large projects such as the one with 19 floors or 485,740 SF.

The RATIO ranges from a low of 0.404 to a high of 0.936, with the mean at 0.681. The $perGSF ranges from a low of $39.47 to a high of $260.66, with the mean computed at $126.16. The distributions for the RATIO and $perGSF data are more normally distributed than the FLOOR and GSF data, and can be approximated as being normal (refer to Figures 5.8 and 5.9).

Table 5.5. Descriptive Statistics of the Variables FLOOR, GSF, Ratio, and $/GSF.

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
FLOOR	132	18	1	19	2.74	2.233	4.226	0.211	25.675	0.419
GSF	132	482,380	3,360	485,740	82,310	79,608	1.982	0.211	5.483	0.419
RATIO	132	0.532	0.404	0.936	0.681	0.092	0.438	0.211	0.308	0.419
$perGSF	132	$221.19	$39.47	$260.66	$126.16	$39.00	0.252	0.211	0.939	0.419

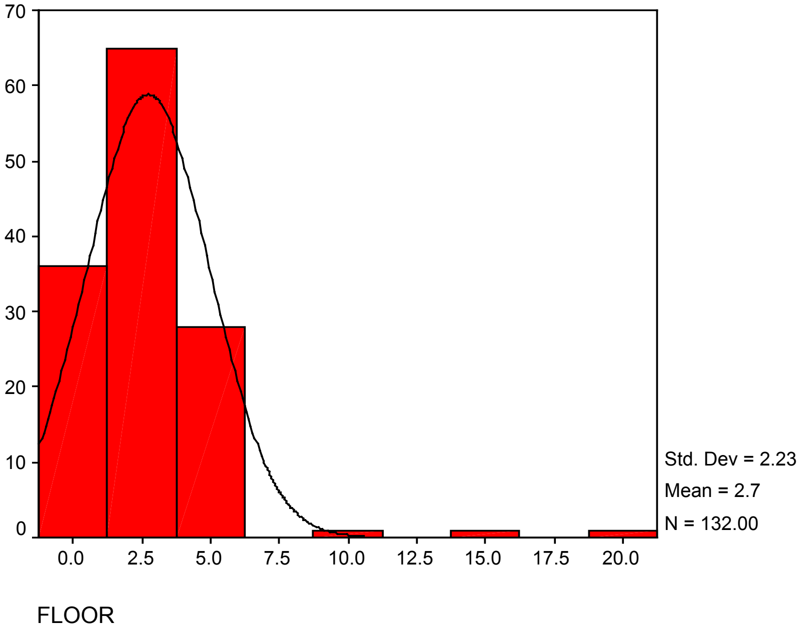


Figure 5.6. Histogram of Number of Floors, with a Normal Curve.

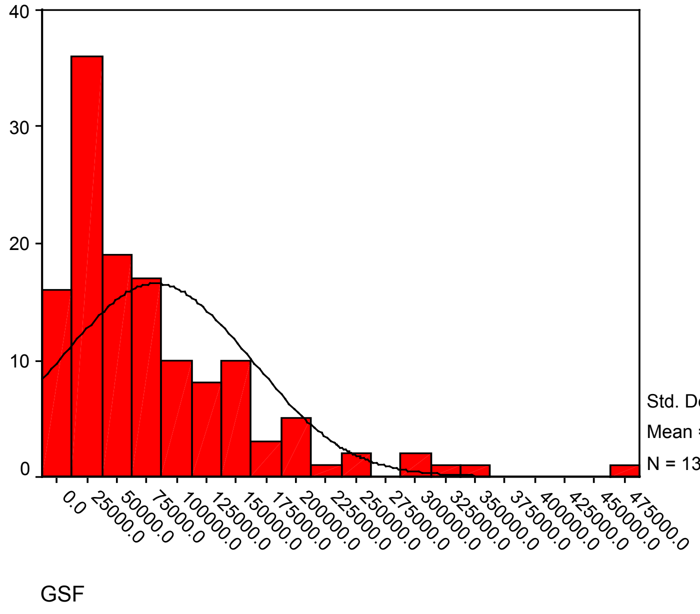


Figure 5.7. Histogram of Gross Square Footage, with a Normal Curve.

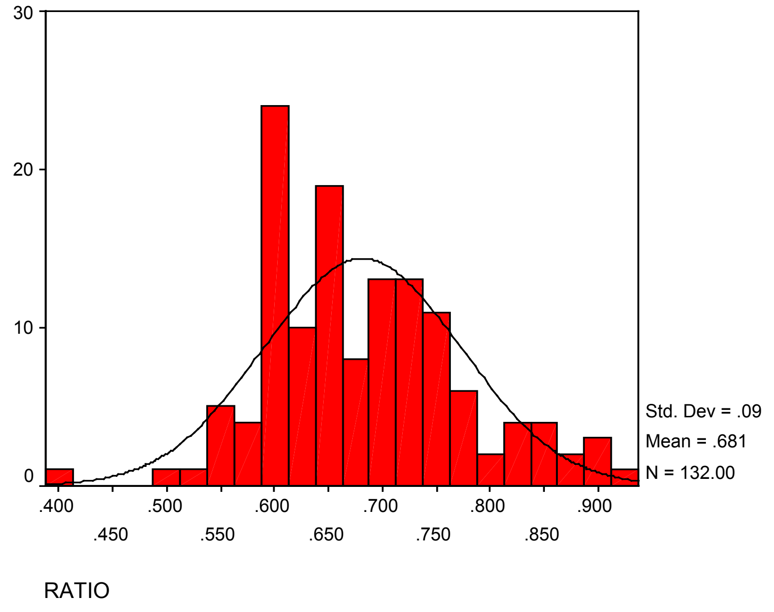


Figure 5.8. Histogram of Usage Ratio, with a Normal Curve.

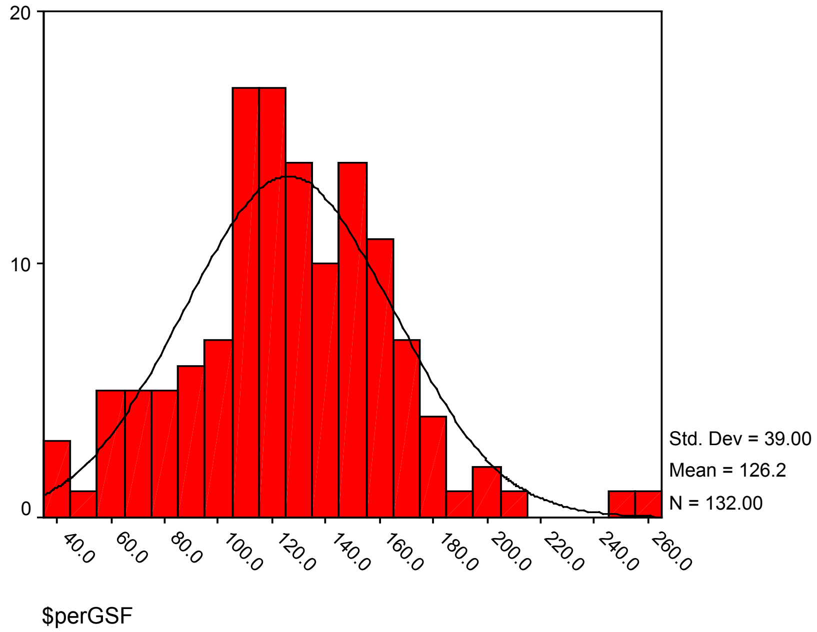


Figure 5.9. Histogram of Building Unit Cost, with a Normal Curve.

5.5.2 Assumptions for Multiple Regression Analysis

Before further discussion on data examination and analysis, it is important to emphasize the assumptions required for multiple regression analysis. Multivariate analysis, such as multiple regression, requires that the assumptions be tested twice: first for the separate variables, and second for the multivariate model variate. A variate is a linear combination of variables formed in the multivariate technique by deriving empirical weights applied to a set of variables specified by the research [Hair et al. 1998].

The most important assumptions for the first tests, for separate variables, are the tests for normality and linearity. Normality of the data distribution for an individual metric variable is the most fundamental assumption in multivariate analysis. If the variation from the normal distribution is sufficiently large, all resulting statistical tests are invalid, as normality is required for the F and t statistics. Linearity is also an assumption for all multivariate analysis based on correlational measures of association, including multiple regression analysis [Hair et al. 1998]. While the first tests are focused on individual variables, the second tests are concerned with the relationship of the variables as a whole. The second tests are carried out after the regression analysis. The assumption to be examined is that the relationship between the dependent and independent variables is linear and that for each combination of values of independent variables, the distribution of the dependent variable is normal with constant variance [Norušis 1995].

5.5.3 Data Transformations

As presented above, the distributions of the FLOOR and GSF variables have strong positive skews. Therefore, incorporating these variables into the model without transformation may seriously violate the normality assumptions. Bowlby and Schriver suggested that the number of floor could be transformed by a natural logarithm as follows:

$$TFLR = \ln(FLOOR)$$

This transformation is also supported by the logic that increasing the number of floors will increase the unit cost at a decreasing rate; that is, increasing the floor by one level, for a one story to two story building, will result in a greater unit rate increase than for a similar one floor increase in a 10-floor building.

The transformation suggested for the GSF is as follows:

$$TGSF = 1 \div (GSF \div 100)$$

Similarly, the rational behind this transformation is that although there are economies of scale for constructing bigger buildings, these economies would be realized very quickly. That is, increasing the floor space from 1000 to 2000 square feet would lower the cost by 75 cents per square foot, while increasing the size from 10,000 to 11,000 would lower the cost by only 1.36 cents [Bowlby et al. 1986].

The descriptive statistics for the transformed variables can be seen in Table 5.6.

Table 5.6. Descriptive Statistics of the Variables TFLR and TGSF.

	N	**Range**	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**	**Skewness**		**Kurtosis**	
	Statistic	**Statistic**	**Statistic**	**Statistic**	**Statistic**	**Statistic**	**Statistic**	**Std. Error**	**Statistic**	**Std. Error**
TFLR	132	2.94	0.0	2.94	0.8080	0.61136	0.323	0.211	0.301	0.419
TGSF	132	0.03	0.0	0.03	0.0035	0.00466	2.913	0.211	10.202	0.419

The skewness and kurtosis statistics are better for the TFLR variable. However, the transformation did not improve the normality of the TGSF variable.

5.5.4 Accessing Normality

The normalities of the variables are accessed through normal probability plots or Q-Q plots. The four plots for TFLR, TGSF, RATIO, and $perGSF can be seen in Figures 5.10, 5.11, 5.12, and 5.13.

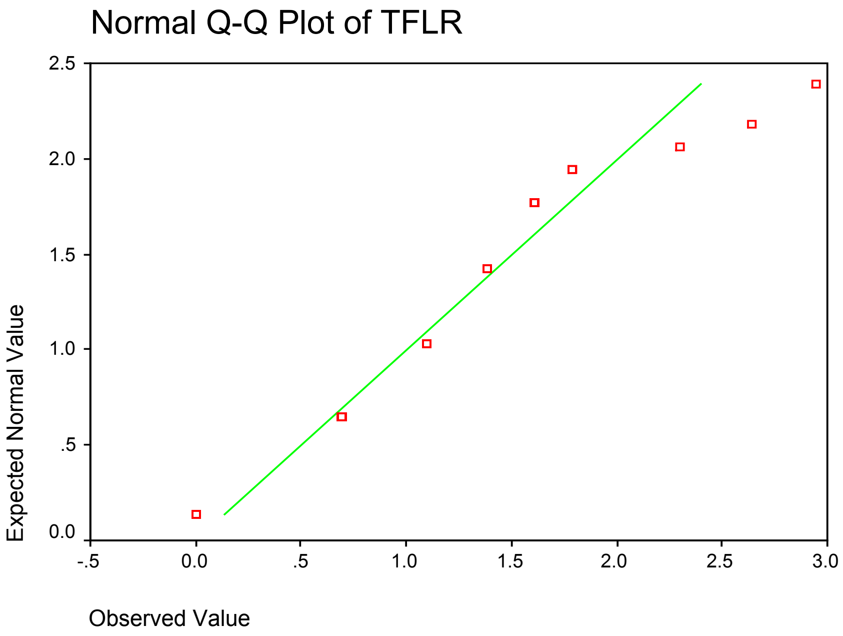


Figure 5.10. Q-Q plot for TFLR.

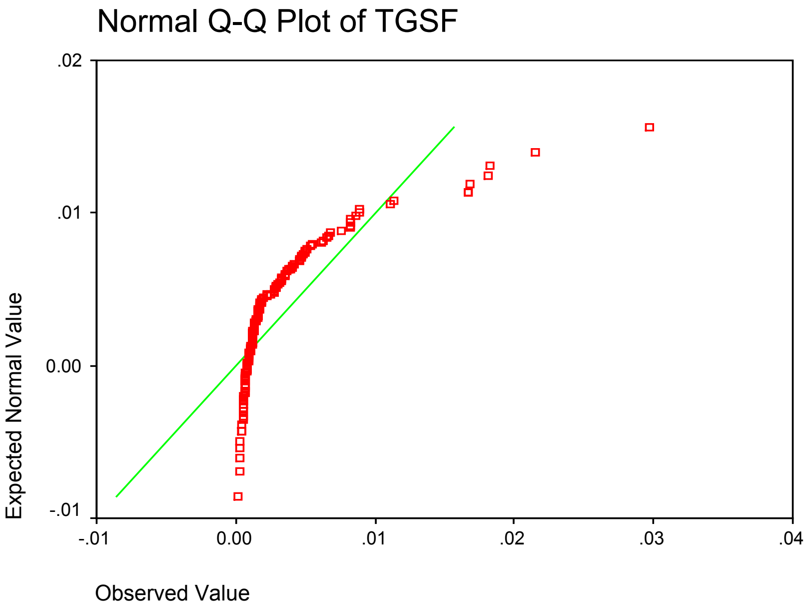


Figure 5.11. Q-Q plot for TGSF.

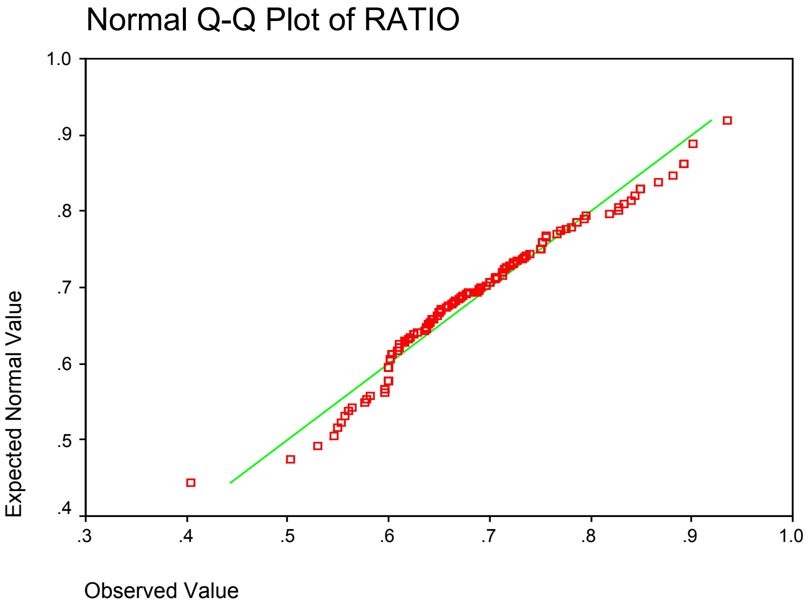


Figure 5.12. Q-Q plot for RATIO.

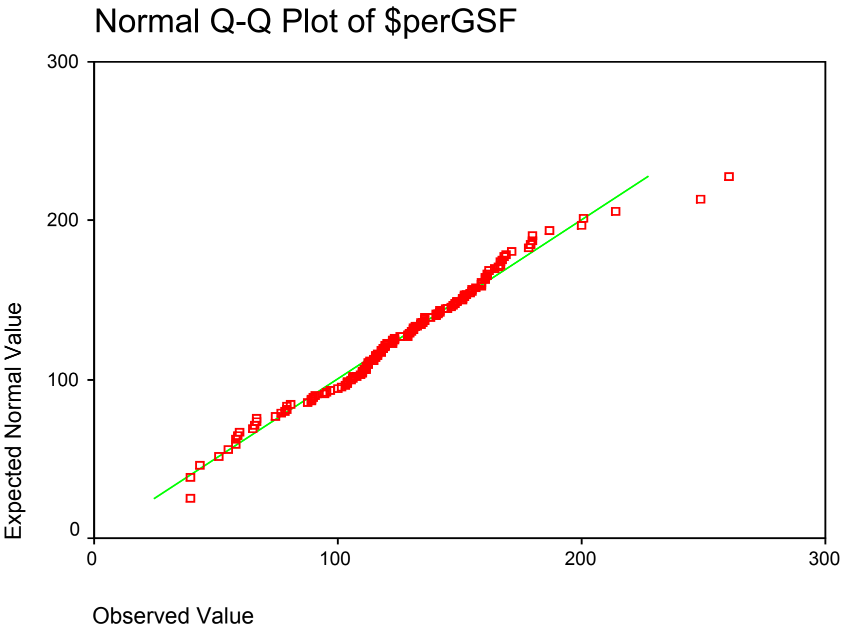


Figure 5.13. Q-Q plot for $perGSF.

For TFLR, RATIO and $perGSF, it can be seen that the data generally cluster around the straight line, with few cases of outliers in the low and high ranges of the observed values. Consequently, it can be concluded that the normal probability plots for TFLR, RATIO and $perGSF are more or less linear, so the assumption of normality appears to be reasonable. However, the distribution of TGSF is still positively skewed as indicated previously in Table 5.5.

5.5.5 Accessing Correlation and Linearity in the Relationships

Before the implementation of the multiple regression analysis, it is important to make sure that the independent variables are linearly related to the dependent variable. The Pearson product-moment correlation is used to assess the zero-order correlations among the variables. The calculated correlation matrix for the independent variables and dependent variable can be seen in Table 5.7, while Figure 5.10 shows the corresponding scatterplot matrix.

Table 5.7. Correlation Matrix For The Independent Variables And Dependent Variable.

Variables	**TFLR**	**TGSF**	**RATIO**	**$perGSF**
TFLR	1.000	-.559	-.506	.546
TGSF		1.000	.297	-.251
RATIO			1.000	-.570
$perGSF				1.000

All correlations are significant at the 0.01 level (2-tailed). (N = 132)

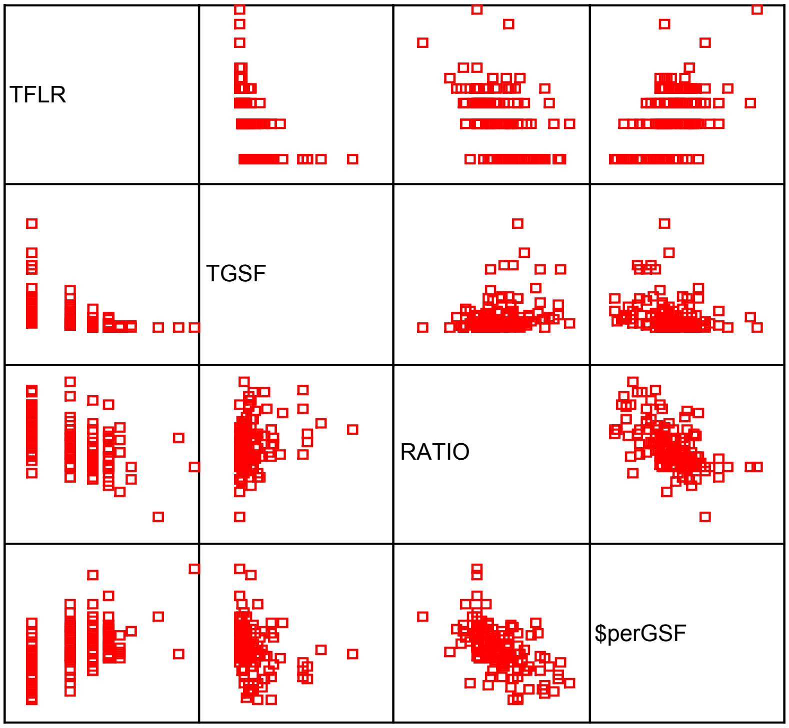


Figure 5.14. Scatterplot Matrix For The Independent Variables And Dependent Variable.

The correlation matrix shows that all the independent variables have significant correlations (at the 0.01 level) with the dependent variable. TFLR has a positive correlation (0.546) with the $perGSF, that is, increasing the number of floors is associated with higher unit cost. TGSF has a weak negative correlation (-0.251) with the $perGSF, indicating that increasing the floor area is associated with lower unit cost. RATIO also has a negative correlation (-0.570) with the $perGSF, indicating that the higher the usage ratio, the lower the unit cost. The correlation of the RATIO and $perGSF is the strongest for the three independent variables considered.

These correlations can also be graphically examined via the scatterplot matrix shown in Figure 5.10. By examining the individual scatterplots on the last row of the matrix, the linearity of the relationships between the independent variables and the dependent variable can be assessed. It can be seen that all the independent variables have linear relationships with the dependent variables. With the above data examinations, data transformations, and the relationships assessment, the pre-analysis for multiple regression analysis is successfully performed. The next section discusses the execution of the multiple regression analysis.

5.6 MULTIPLE REGRESSION ANALYSIS

This section discusses the multiple regression analysis methodology used in this study for the parametric cost estimating model development. The section begins by first discussing the model. The detailed steps in the analysis are then presented and explained, which is followed by the results presentation. Lastly, the diagnostics and assumption verifications of the multiple regression analysis are discussed.

5.6.1 Model Description

The technique of multiple regression analysis is employed in this study to determine the mathematical function that relates the building parameters to the unit cost of the building. The result of the multiple regression analysis is this mathematical function, sometimes referred to as the multiple regression equation. This multiple regression equation is our parametric cost model. This is the heart of the parametric cost estimating method. The form of the equation will be as follows:

$$\hat{Y} = b_0 + b_1 V_1 + b_2 V_2 + b_3 V_3$$

where $\hat{Y}$ = Conceptual cost estimate (\$perGSF)
b_0 = Regression constant
b_1 = Partial regression coefficient of V_1
V_1 = TGSF
b_2 = Partial regression coefficient of V_2
V_2 = TFLR
b_3 = Partial regression coefficient of V_3
V_3 = RATIO

Multiple regression analysis is used as a methodology for determining the *b* values, the constant and partial regression coefficients in the equation. In addition, statistical significances of the variables as predictors are also assessed. The goal of multiple regression analysis is to build a simple model that predicts well. If the model predicts equally well with two variables instead of five, the simpler model is always better. This is the concept of parsimony, which refers to the balance of simplicity and fit of the model.

5.6.2 Model Development

In building the model, the three variables are entered hierarchically into the equation. The decision was made to enter the variables in the order of increasing bivariate correlations with the dependent variable. As such, TGSF is entered first, followed by TFLR and RATIO.

The results of this first run can be seen in Table 5.8. It is evident that all the models are significant and the final R^2 for the model with the three independent variables is 0.420. In addition, although TGSF when considered alone in the equation is significant (t = -2.954), the corresponding model has a low R^2 of 0.063. TGSF also became insignificant when another independent variable is added into the equation; the TGSF's t = 0.899 when TFLR is added and TGSF's t = 1.079 when RATIO is also added.

Table 5.8. Multiple Regression Run #01: TGSF, TFLR and RATIO versus $perGSF (N = 132).

Variables in Regression Equation	**R** *****	**R^2** ******	**Adjusted R^2** *******	**F** **(df)**	**Sig.**
TGSF	0.251	0.063	0.056	8.726 (1, 130)	0.004
TGSF, TFLR	0.550	0.303	0.292	28.006 (2, 129)	0.000
TGSF, TFLR, RATIO	0.648	0.420	0.406	30.867 (3, 128)	0.000

---

*R, the multiple correlation coefficient, is the correlation between the observed and predicted values of the dependent variable. The values of R for models produced by the regression procedure range from 0 to 1. Larger values of R indicate stronger relationships.

**R^2 is the proportion of variation in the dependent variable explained by the regression model. The values of R^2 range from 0 to 1. Small values indicate that the model does not fit the data well. R^2 is generally used to determine which model is best. The best model is one with a high value of R^2 that does not contain too many variables.

***Adjusted R^2 attempts to correct R^2 to more closely reflect the goodness of fit of the model in the population. The sample R^2 tends to optimistically estimate how well the models fits the population.

Any serious multivariate outlier for this study is detected through a casewise diagnostics report. A casewise diagnostics report displays the statistics for cases with standardized residuals more than a specified number of standard deviations away from the mean. Standardized residuals are ordinary residuals divided by the sample standard deviation of the residuals. Standardized residuals have a mean of 0 and a standard deviation of 1 [Norušis 1995]. For this analysis, 3 standard deviations are specified, and one case is identified with standard residuals of 3.391. The project is the Radiation Oncology Center (Project ID 136).

To improve the model, the Radiation Oncology Center is removed and the analysis repeated. The result can be seen in Table 5.9.

Table 5.9. Multiple Regression Run #02: TGSF, TFLR and RATIO versus $perGSF (N = 131).

Variables in Regression Equation	**R**	**R^2**	**Adjusted R^2**	**F (df)**	**Sig.**
TGSF	0.260	0.068	0.060	9.354 (1, 129)	0.003
TGSF, TFLR	0.560	0.314	0.303	29.291 (2, 128)	0.000
TGSF, TFLR, RATIO	0.655	0.429	0.416	31.817 (3, 127)	0.000

All the models show slight improvements with the removal of the Radiation Oncology Center. The final R^2 for the model with the three independent variables is 0.429. However, TGSF still remains insignificant when another independent variable is added into the equation; the TGSF's t = 0.852 when TFLR is added and TGSF's t = 1.036 when RATIO is also added.

In consideration for model parsimony, the regression analysis is now performed without the TGSF. The results are presented in Table 5.10.

Table 5.10. Multiple Regression Run #03: TFLR and RATIO versus $perGSF (N = 131).

Variables in Regression Equation	**R**	**R^2**	**Adjusted R^2**	**F (df)**	**Sig.**
TFLR	0.557	0.310	0.305	57.979 (1, 129)	0.000
TFLR, RATIO	0.651	0.424	0.415	47.163 (2, 128)	0.000

It is evident that dropping the TGSF has very little impact on the model developed. For the model with only TFLR and TGSF, removing TGSF changes the R^2 only slightly from 0.314 to 0.310. Similarly, for the full model with all the three independent variables, dropping the TGSF also changes the R^2 only slightly from 0.429 to 0.424.

Table 5.11 presents all the equations derived from the three regression analysis runs. All the unstandardized coefficients are listed along with the associated t-values in parentheses. The F value and the R^2 for each equation are also provided. The number of projects for Run 1 is 132. Run 2 and 3 are conducted after the removal of the Project 136 Radiation Oncology Center, and thus are based on 131 projects.

Table 5.11. All Regression Equations Developed Through Multiple Regression Analysis.

Eqn.	Constant	Independent Variable			F	R^2
		TGSF	TFLR	RATIO		
Run1 A	133.516 (32.300)	-2099.914 (-2.954)			8.726	0.063
Run1 B	93.363 (13.318)	667.570 (0.899)	37.687 (6.661)		28.006	0.303
Run1 C	218.296 (8.590)	734.038 (1.079)	25.154 (4.383)	-168.862 (-5.081)	30.867	0.420
Run2 A	132.555 (33.235)	-2092.455 (-3.059)			9.354	0.068
Run2 B	93.478 (13.933)	605.784 (0.852)	36.753 (6.780)		29.291	0.314
Run2 C	212.823 (8.733)	674.592 (1.036)	24.859 (4.526)	-161.323 (-5.060)	31.817	0.429
Run3 A	97.686 (21.530)		34.167 (7.614)		57.979	0.310
Run3 B	216.997 (9.025)		22.032 (4.621)	-160.633 (-5.038)	47.163	0.424

Dependent Variable: \$perGSF
t-values shown in parenthesis
N= 132 for Run1, and N = 131 for Run2 and Run3

It is evident that for similar model structure, the coefficient values have only slight changes for the various runs. In selecting the best model, the R^2 for each equation is examined. A high R^2 indicates the model has the best fit for the data and is generally preferred. Equation Run2 C has the highest R^2, but the TGSF's t-value is not significant. Alternatively, equation Run3 B has the next highest R^2 value, slightly less than that of Run2 C; therefore considering model parsimony, equation Run3 B is selected as the final model for predicting the $perGSF.

5.6.3 Multiple Regression Diagnostics

An important part of regression analysis is checking that the required assumptions are met. Residual analysis was performed to evaluate the assumptions of linearity, normality and homoscedasticity. The residuals are the differences between the observed and the predicted values. Homoscedasticity is a description of data for which the variance of the error terms appears constant over the range of values of an independent variable [Hair et al. 1998].

First, a histogram of the standardized residuals and normal probability plot was run to check the linearity and normality assumption [Norušis 1995]. The histogram and the normal probability plot can be seen in Figures 5.15 and 5.16, respectively. In addition, the statistical test of normality for the standardized residuals indicated that the normality assumption was not violated (Kolmogorov-Smirnov = 0.074, df = 131, p = 0.074).

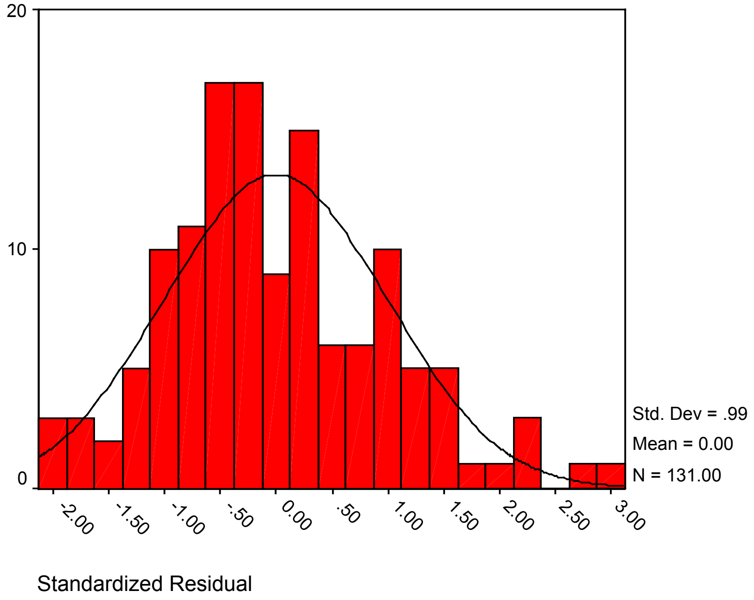


Figure 5.15. Histogram of the Standardized Residuals (N=131).

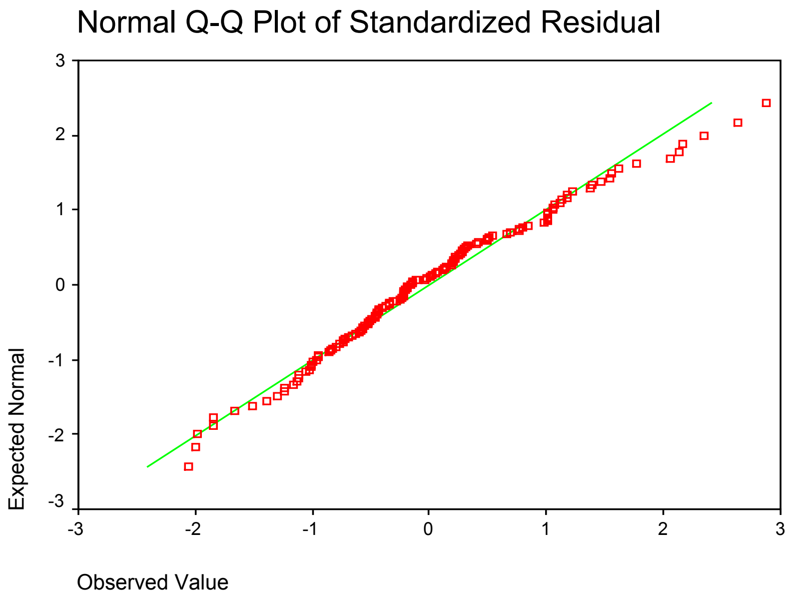


Figure 5.16. Normal Probability Plot for the Standardized Residuals (N=131).

Second, the residuals were plotted against the predicted values (Figure 5.17) to test for the homoscedasticity assumption. The plots of residuals appear to be randomly scattered around a horizontal line; therefore, the assumption of constant variance and linearity has been met.

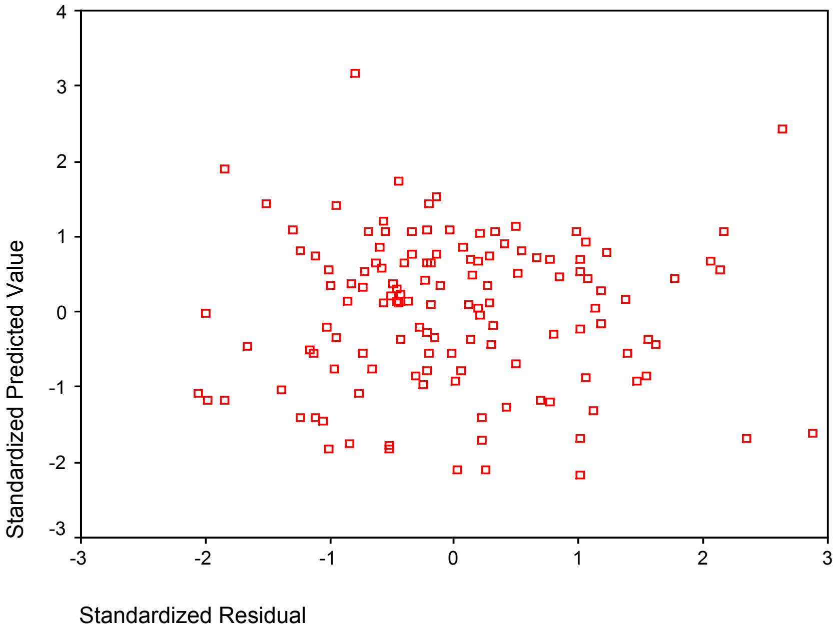


Figure 5.17. Scatterplot of Standardized Residual versus Standardized Predicted Value (N=131).

Finally, the multicollinearity issue of the model is assessed. Multicollinearity exists when independent variables or predictors are highly correlated with each other, which can cause logical and statistical difficulties. With multicollinearity, the estimation of weighting coefficients is unstable. To assess multicollinearity among the variables, tolerances and variance inflation factors (VIF) were examined.

Tolerance refers to the proportion of the variance of that variable that is not accounted for by other predictors in the model. The range of tolerance is from 0 (perfect collinearity) to 1 (no collinearity). A tolerance with values less than 0.1 typically indicates the multicollinearity problem [Norušis 1995].

Variance inflation factor (VIF) is another index for the diagnostic of multicollinearity. The high value of a variance inflation factor for a variable indicates there is a strong association between that variable and other remaining predictors [Stevens 1996]. If a variable has high tolerance, it will have a small variance inflation factor. A variance inflation factor in excess of 10 indicates the multicollinearity problem [Stevens 1996].

Since our final model has only two predictors, both the tolerances are calculated to be 0.745 and both the VIF values are 1.343. Therefore, multicollinearity was not a serious problem for this analysis.

5.6.4 The Derived Cost Estimating Model

The equation selected from the previous discussion, Section 5.6.2, becomes the parametric cost estimating model function. The equation or the model function can be written as:

\$perGSF = 216.997 + 22.032 TFLR – 160.633 RATIO

or

Cost estimate in \$ per GSF

= 216.997 + 22.032 ln(number of floor) – 160.633 (usage ratio) *(EQ 5.1)*

The equation's R^2 value is 0.424 indicating that 42.4% of the variation in the \$perGSF (dependent variable) is explained or accounted for by the variations in the TFLR and RATIO (independent variables). The remaining 57.6% of the variations is due to other factors not included in the model, such as the building type, framing system, and the type of institutions submitting the projects.

In addition, the relative importance of the independent variables, TFLR and RATIO, can be assessed by examining their respective standardized coefficients. Predictors with higher standardized coefficients are more important to the regression equation than those with lower values [Tabachnick and Fidell 2001]. Examining the associated standardized coefficients for the TFLR (0.359) and RATIO (-0.392), the RATIO with higher standardized coefficient magnitude can be inferred as the more significant and more powerful predictor for the dependent variable, $perGSF.

The model's prediction performance based on the data used in its development can be summarized in Figure 5.18. It can be seen that the model performs quite well with most of the predictions. 48 out of 131 projects (36.6%) have prediction values within the 10% deviations from the observed value. For our discussion, the percent deviation is determined by Predicted $/GSF minus Observed $/GSF divided by Observed $/GSF. Less than 8% of the projects have more than 50% deviation.

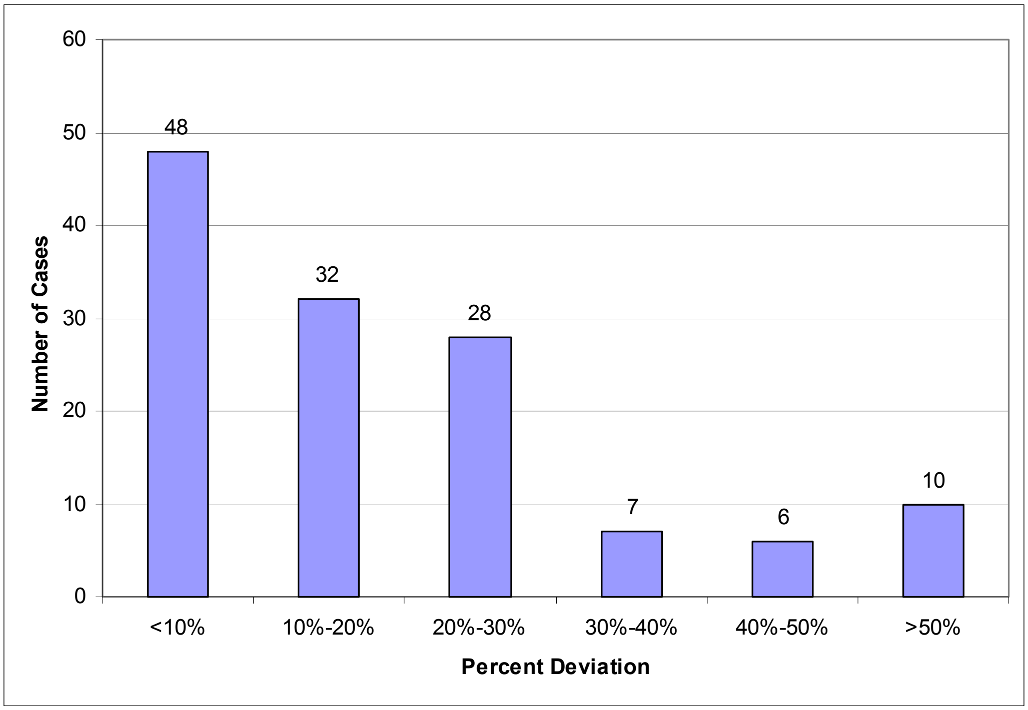


Figure 5.18. Histogram of the Model Prediction Performance. (N = 131)

However, with twenty-three projects having cost deviations of more then 30%, additional analyses are deemed necessary to investigate these more extreme cases of cost deviations. The description of the twenty-three projects can be seen in Table 5.12

Table 5.12. Description of Project Cases with More Than 30% Cost Deviations

30%-40%

	ID	Name	Type	TSTC	Cost Deviation
1	162	Medical Education Division	Classroom		-31%
2	4	Athletic Facility	Athletic Facility		31%
3	75	University Interscholastic League Building	Administrative/Office		32%
4	40	Academic Building	Classroom		32%
5	77	Computer Application Center	Classroom	Yes	33%
6	90	Good Lab Practices Facility	Research Laboratory		35%
7	138	Faculty Center	Administrative/Office		37%

40%-50%

	ID	Name	Type	TSTC	Cost Deviation
1	157	Student Health Center	Health Care		-49%
2	147	Coastal Engineering Laboratory	Research Laboratory		-45%
3	76	Science and Technology Building	Classroom	Yes	41%
4	20	Aerospace/Industrial Building	Classroom	Yes	41%
5	12	Companion Animal Geriatric Center	Administrative/Office		42%
6	61	Academic/Student Service	Classroom		45%

>50%

	ID	Name	Type	TSTC	Cost Deviation
1	7	Engineering Graphics Building	Classroom	Yes	52%
2	33	Activity Center	Student Center	Yes	55%
3	78	Fentress Center	Classroom	Yes	56%
4	62	Recreational Sports	Athletic Facility	Yes	65%
5	52	District Headquarter	Administrative/Office		67%
6	101	Communication Disorder/Psychology Building	Administrative/Office		72%
7	176	Theriogenology Facility	Research Laboratory		86%
8	24	Aviation Technology Building	Classroom	Yes	123%
9	22	Transition Building	Administrative/Office		145%
10	13	Applied Technology Education Center	Classroom	Yes	149%

It is evident that many of the projects are Texas State Technical College projects (indicated in the table under TSTC column). Of all the thirteen projects built by Texas State Technical College, nine projects are predicted poorly with cost deviations of more than 30%. The projects by Texas State Technical College tend to be over predicted due to the fact that these projects are usually more economical, low-cost facilities when compared to those facilities built by major universities. Another explanation is that the technical college generally have less funding to construct their facilities then the major universities, thus their facilities are not build to the higher standard as are those constructed by the major universities such as the University of Texas at Austin. As a result, these Texas State Technical College projects are determined to be different from the general population of the projects considered in this study, and consequently the decision was made to exclude the Texas State Technical College from this study.

In addition, to further refine the model, decision was made to remove the three high-rise buildings as listed in Table 5.13. These three high-rise projects are determined to be different from the general buildings population considered in the study, which have seven floors or less, and their inclusion may not be representative of the types of projects considered. In this way, the model produced can be better defined to estimate projects having seven floors or less. Finally, the transition building (ID 22) was also identified as a temporary facility project, thus having poor cost predictions. As a result it was also removed from the data set.

Table 5.13. Project Cases with More Than Seven Floors

ID	Name	Type	# of Floors
86	North Campus - Phase III	Health Care	10
138	Faculty Center	Administrative/Office	14
185	Basic Sciences Research Building Phase II	Research Laboratory	19

With the new project data file, whereby the technical college projects were removed, along with the three high rise building projects and the transition building, addition analysis was performed on the data. Multiple regression analysis technique was again applied and the new model developed as follow:

$$\$perGSF = 202.245 + 15.740\ TFLR - 126.196\ RATIO$$

or

$$\text{Cost estimate in \$ per GSF}$$

$$= 202.245 + 15.740 \ln(\text{number of floor}) - 126.196\ (\text{usage ratio}) \qquad \text{(EQ 5.2)}$$

The equation's R^2 value is 0.261 indicating that 26.1% of the variation in the \$perGSF (dependent variable) is explained or accounted for by the variations in the TFLR and RATIO (independent variables). The remaining 73.9% of the variations is due to other factors not considered in the model development.

The new model's prediction performance, based on the data used in its development, can be summarized in Figure 5.19. It is evident that the model's performance improves. The proportion of the projects having less then 20% cost deviation increases from 61% to 72%. The number of projects with more then 30% cost deviation also dropped from 23 to 16 projects, with only 3.5% of the cases having more then 50% cost deviations.

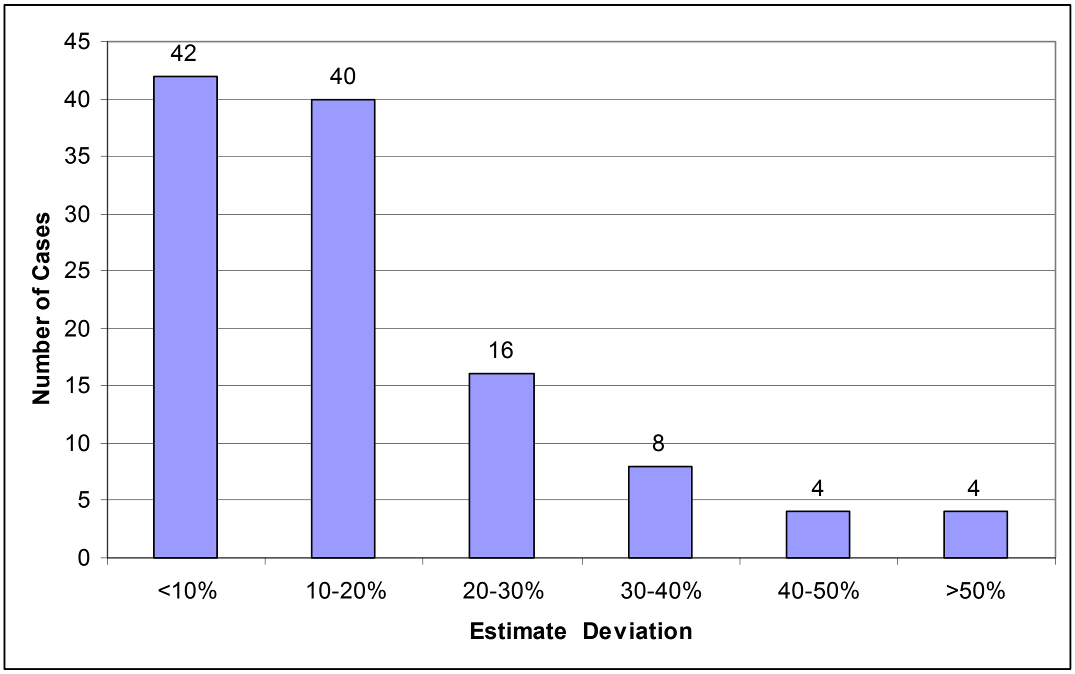


Figure 5.19. Histogram of the Model Prediction Performance. (N = 114)

Effort was also made to include a new variable into the data in order to further classify the projects. Building projects can be classified as ordinary building, structural intensive building, or mechanical intensive building. Structural intensive buildings are associated with parking garage building type and stadium building type, where most of the construction costs are related to the structural constructions. Mechanical intensive buildings are associated with laboratory and dormitory building types where substantial amount of the project costs are related to the costs of the mechanical systems and plumbing systems. Since the data set do not contained parking garage building type and stadium, the projects in the data set are only classified as mechanical intensive buildings or ordinary buildings. The new classification variable created is named “DORMLAB”, and takes a value of either a “1” or “0”. The projects in the data set are coded “1” if the project type is research laboratory or dormitory, and coded “0” for all other building types. Multiple regression analysis was then performed and the resulting model developed as shown below:

$$\$perGSF = 201.327 - 5.442\ DORMLAB + 15.540\ TFLR - 123.860\ RATIO$$

or

$$\begin{aligned} &\textit{Cost estimate in \$ per GSF} \\ &= 201.327 - 5.442\ \textit{DORMLAB} + 15.540\ \ln(\textit{number of floor}) \\ &\quad - 123.860(\textit{usage ratio}) \end{aligned} \qquad (EQ\ 5.3)$$

Examining the coefficients in the equation, it is evident that the derived model is not significantly different from the previous model (*EQ 5.2*). In addition, the variable DORMLAB is also not statically significant at the .05 level (P = 0.539). The results of the predictions also produced no improvement, therefore this new model was not accepted and final model adopted is the previous model as defined by Equation 5.2.

With the model developed, the next essential step is the validation of the model. The discussion of the model validation is presented in the next section.

5.7 MODEL VALIDATION

One of the most important steps in developing a cost model is to test its accuracy and validity [Dysert 2001]. This process is also refers to as the model validation; it involves the testing and evaluating of the developed model with some test or validation data. The validation data should be some representative data from the target population. These data are not used in the model development, but are used specifically for testing the developed model. For this study, the validation data is extracted from the project historical data file as briefly discussed in Section 5.4.4. The following subsections elaborate this model validation process by first discussing the details of the validation data set. The validation methodology is then presented. Finally, the discussions on the validation results are provided.

5.7.1 Validation Data Set

The validation data for this study consists of seven projects. The projects are randomly extracted from the historical data file and kept for this validation process, while the remaining 132 projects are used in the model development. The details of the seven projects can be seen in Table 5.14.

Table 5.14. Description of the Validation Data Set.

Project ID	Project Name	Facility Type	Number of Floor	GSF	ASF	Usage Ratio
50	University Services Center	Administrative/Office	2	21,284	14,713	0.691
43	Special Events Center	Multi-Purpose	2	230,000	177,821	0.773
2	Health Science Building	Classroom	4	106,350	74,445	0.700
23	Allied Health Technology Building	Classroom	1	33,743	25,371	0.752
48	Early Childhood Center	Classroom	2	53,416	31,378	0.587
72	Student Health Center	Healthcare	1	6,800	5,508	0.810
26	Student Development Center	Student Center	3	100,760	67,625	0.671

The seven projects represent five university facility types. Most of the buildings are low-rise to mid-rise, with the number of floors ranging from one to four floors. The gross square footage ranges from 6,800 SF to 230,000 SF, and the usage ratio ranges from 0.587 to 0.810.

However, decision was later made to remove the Texas State Technical College from the data set and Project ID 23, Allied Health Technology Building, was identified as technical college project, therefore it is removed from the data set. The remaining six projects are then used for the model validation.

5.7.2 Model Validation Methodology

After the model is developed, to use the model to calculate an estimate involved merely the substitution of the new building parameter values into the equation to calculate the estimated cost. An example of this computation can be seen below for the first case in Table 5.14, University Service Center:

Cost estimate in \$ per GSF

= 202.245 + 15.740 ln(number of floor) – 126.196 (usage ratio)

= 202.245 + 15.740 ln(2) – 126.196 (0.691)

= 125.95

The predicted cost computed is then compared to the observed costs recorded. The next section presents the results of these validations.

5.7.3 Discussion of Validation Results

Based on the methodology above, the predicted costs for all the cases in the validation data set can be computed. The results of the computations can be seen in Table 5.15.

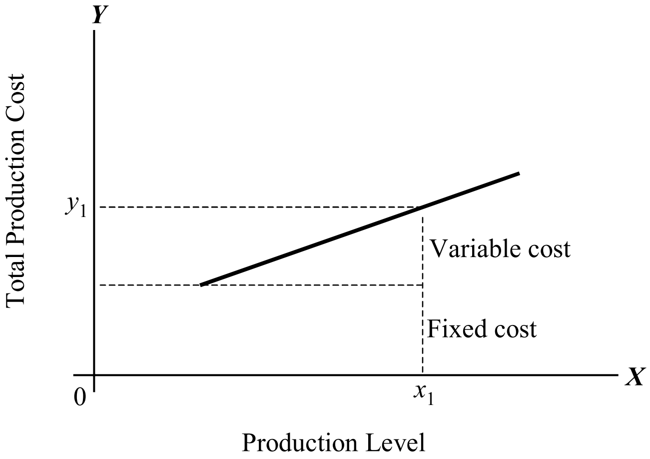
Table 5.15. Comparison of Predicted \$perGSF and Observed \$perGSF for Validation Data Set.

Project ID	Project Name	Facility Type	$ per GSF		%
			Observed	Predicted	Deviation
50	University Service Center	Administrative/Office	125.35	125.95	0.5%
43	Special Event Center	Multi-purpose	115.46	115.61	0.1%
2	Health Science Building	Classroom	136.88	135.73	-0.8%
48	Early Childhood Center	Classroom	137.66	139.08	1.0%
72	Student Health Center	Healthcare	88.11	100.03	13.5%
26	Student Development Center	Student Center	134.45	134.86	0.3%

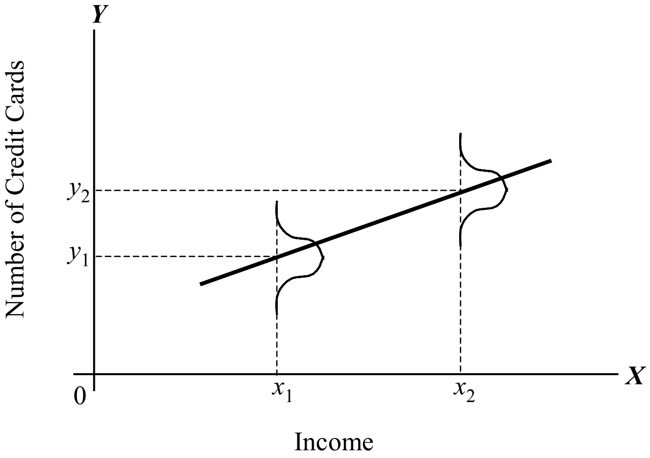
% Deviation = Predicted \$/GSF minus Observed \$/GSF divide by Observed \$/GSF

It is evident that the model performed very well with the validation data set, with predictions ranging from underestimating by –0.8% to overestimating 13.5%.

In order to enhance the understanding the model and its validity, it must be highlighted that the parametric model developed in this study is based on statistical relationships. In a statistical relationship, more than one value of the dependent values will usually be observed for any value of an independent variable. The model developed establishes the statistical relationship, which unlike a functional relationship, estimates an average value instead of an exact value. Both of these relationships are shown in Figure 5.20.



(a) **Functional Relationship**



(b) **Statistical Relationship**

Figure 5.20. Comparison of Functional and Statistical Relationships (Adapted from [Hair et al. 1998])

As such, the predicted cost will never be exactly equal to the actual costs. In a statistical study, there will always be some random components to the relationships being examined. This observation is important in the understanding of the accuracy and value assessment of the estimate produced by the statistical model.

5.8 LIMITATION OF THE ANALYSIS

As in all studies, the limitations of the analysis must be noted and considered along with the results presented. This section discusses the limitations as applied to the data analysis performed in this study. The limitations can be classified into two types: data related issues and methodology related issues.

5.8.1 Data Related Issues

The first data related issue is that concerning data integrity and accuracy. Since the data for this study is retrieved from the previously submitted THECB Construction Application Form, and that no additional information regarding the subsequent project changes and the actual costs incurred is available, the THECB data is not verified. However, it can be said that the project costs should not deviate by more than 10 percent, since greater than 10 percent difference would require resubmission of the construction application.

The second issue concerns the domain of the project types considered in this study. The data used in the analysis are very focused and domain specific, relating to university-type, high occupancy, public building projects constructed in Texas. Therefore, generalizing the model developed to other project type, or to other organization sector, or to other locality may not be readily feasible. Although the methodology used in this analysis can be adapted to perform similar analysis based on relevant data from that other project domain.

The third issue relates to the model implementation. Due to the fact that the model is developed from the limited historical project data collected, the usage of the developed model is limited to application to the proposed projects that are in the range of the data used to develop the model. Using the model to predict the cost of the future projects that are not representative of the data use in the model development is an extrapolation, which may produce unreliable predictions.

5.8.2 Methodology Related Issues

The first issue is related to the data collection procedure. This research uses the unobtrusive research method, whereby existing data is collected for use in the analysis. In this manner, the researcher has less control over the data contents, quality and quantity. Consequently, the data analysis and model development is highly influenced by these available data.

The second issue is that the model is developed based on the past projects. In this way, the development of the model and implementation of the model is carried out with the assumption that the future predictions is a function of the past trends and pattern of occurrences. Therefore, to maintain and improve the validity of the model, the model must be continually updated to reflect the changes in construction costs and methodologies that may affect the variations in the project costs.

The third issue is that the model developed is based on the analysis of various institutional building types, such as offices and classrooms. Although it has been suggested that the performance of the model developed for a more homogeneous building type is better, due to the limited data available this research uses the data that include various institutional building types. The model developed can generalize to this various institutional building types, but it is anticipated that a better model can be developed when data become available permitting separate analysis for each building type.

5.9 CHAPTER SUMMARY

This chapter presents comprehensive discussions on the data analysis and model development. Various statistical techniques are used in the data examinations and analyses. Specifically, multiple regression analysis is used in the parametric cost estimating model development.

Three building parameters are initially selected as the independent variables for predicting the single dependent variable, building cost in dollars per gross square foot. The three parameters are the gross square footage, number of floors, and the building usage ratio. The gross square footage and the number of floors are the scale related building parameters. The usage ratio, which is the ratio between the building assignable square footage and gross square footage, is proposed as a quality related measure variable for a building construction project. To enhance the logic of the developed model and to improve the analysis, the gross square footage and number of floors variables are transformed. Subsequent multiple regression analysis concludes that the only significant variables for the parametric cost estimating model are the number of floors and the building usage ratio. The building usage ratio is also identified as the more powerful predictor for estimating the building unit cost.

From the model development discussions and validation results, it can be concluded that the developed model performed satisfactorily. For the model development data set, the percent cost deviations are quite low with about 72% of the project cases with prediction deviations of less than 20%, and only 3.5% of the data have prediction deviations of more than 50%. For the validation data set, the results of the predictions for all the cases are within the 15% deviations. In addition, the model is statistically significant at the .05 level (F = 19.624, df = 2, 111) with $R^2 = 0.261$ and all the regression assumptions have been satisfied.

Therefore, the model can be considered a reliable tool for performing the conceptual cost estimating.

With the parametric cost estimating model developed, the next chapter presents the computerized approach to the conceptual cost estimating based on this derived model.

CHAPTER 6
COMPUTERIZED APPROACH OF THE PARAMETRIC COST ESTIMATING MODEL

6.1 INTRODUCTION

This chapter presents the implementation of a microcomputer to automatically generate a conceptual cost estimate. A computer program called the Parametric Cost Estimating Model (PCEM) is developed based on the parametric cost estimating function derived as presented in the Chapter 5. First, the needs and benefits of the computerized approached are discussed. Next, the description of the PCEM computer program are presented, which is followed by a step-by-step demonstration of the program execution.

6.2 NEEDS AND BENEFITS OF A COMPUTERIZED APPROACH TO CONCEPTUAL COST ESTIMATING

Unlike detailed cost estimating, which is mostly a systematic process of estimate generation whereby known quantities are multiplied by known unit costs and totaled; generally conceptual cost estimate lacks such standardized procedure for estimate execution. There is a need for a standardized and systematic approach to conceptual cost estimating so as to improve the conceptual cost estimating processes and lending consistency to the estimates produced. Standardizing a process is one of the first steps to systematic benchmarking and improving a process. A systematic approach to conceptual cost estimating also provides framework for a quantitative method of estimate generation, thereby minimizing the inherent subjectivity in the conceptual cost estimating process.

To address these needs, detailed steps to conceptual cost estimating are first identified and a computerized approach are developed and implemented to facilitate this standardization effort. Utilizing PCEM computerized approach has a myriad of benefits; such as:

a. Clearly define inputs, outputs, and estimating procedures
b. Automated data retrievals and calculations
c. Error free computations
d. Quick and efficient production of cost estimates
e. Enhanced accuracy and validity
f. Consistent estimate generation
g. Facilitate future project data collection

Based on these needs and potential benefits, the PCEM computer program is envisioned and developed. The next section discusses the detailed description of the program.

6.3 DESCRIPTION OF THE PCEM COMPUTER PROGRAM

This section is divided into three subsections: the first subsection discusses the PCEM program development, the second subsection presents the overview of the PCEM program, and the third subsection discusses the limitation of the PCEM program.

6.3.1 PCEM program development

The PCEM computer program was developed under an IBM® PC environment. The operating system for the computer was Windows® 2000. The software used to develop the application was Microsoft® Visual Basic® Programming System Version 6.0. Visual Basic® Programming System is a tool that allows user to create software applications for the Windows® operating system.

The application developed has a graphical user interface (GUI). This user interface is the way that program accepts instruction from user and presents results. GUI facilitates application usage by the easy intuitive interfaces, which provides visual clues such as clearly defined input area, message prompt for missing values, and button selections for program executions. GUI interface helps the user give instructions to the computer and runs the application.

In addition to the program, two data files are also developed: cityindex.txt and CostUniF.txt. The two data files are text files. The cityindex.txt file provides a list of the cities to be presented to the user for selection. The file also includes the corresponding city indexes for cost estimate adjustment to the selected city. The base city for the index is Houston, with value of 1. The CostUniF.txt file provides a list of the facility types a user can choose from. Associated with each facility type are the ratios of the nine system costs to the total estimated cost. The nine system cost ratios are based on the UniFormat Cost Classification System. The ratios are used to breakdown the total estimated cost into the nine UniFormat Cost Classification System based on the facility type selected. The city cost indexes have been adapted from *Means Building Construction Cost Data 2002*, while *Means Square Foot Cost 2002* was used to derive the UniFormat Cost Classification System ratios.

The PCEM program, and the two data files are to be stored on the system hard drive (C drive) under the folder PCEM. The two data files can be modified, by adding or removing items on the list or changing the numbers to reflect changes, but the format presented should be retained to enable data to be read properly into the program.

The codes for the PCEM application and the contents of two text files are provided in the Appendices.

6.3.2 Overview of the PCEM program

The overview of the PCEM program is illustrated in Figure 6.1. The estimator uses the PCEM program by entering the required inputs into the program. The program uses two data files in its execution and estimate calculations. The output of the program is the cost estimate report.

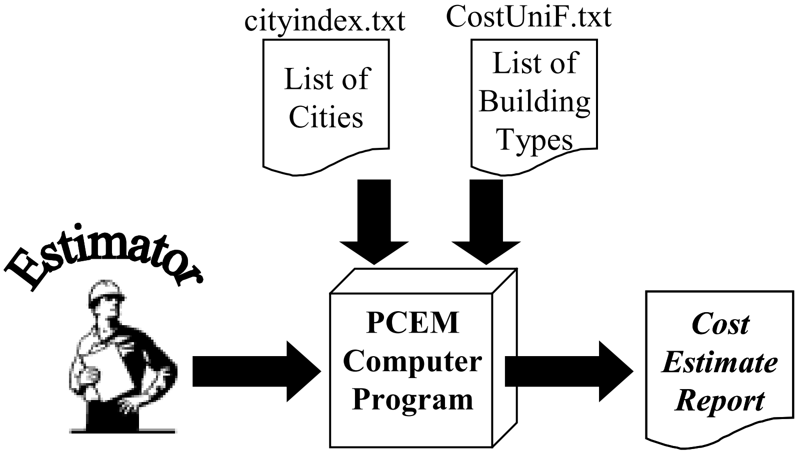


Figure 6.1. Overview of the PCEM Computer Program.

6.3.3 Limitation of Computer Program

The computer program was developed based on the parametric cost estimating function derived in Chapter 5. The scope and limitation to the computer program, in addition to those related to the model development presented in Chapter 5 are as follows:

a. All the formula for all the calculations involved are static and are built into the program. However, these formula can be modified by changing the codes and repackaging the application.

b. Since this program is developed as a preliminary tool to demonstrate the parametric method in conceptual cost estimating. There is no provision for automatic updating of the parametric cost estimating regression equation when new project data becomes available. The new parametric cost estimating equation must be derived using the similar analysis methodology and multiple regression analysis technique as presented in Chapter 5.

c. The program calculates the costs based on the 2001 dollar.

d. The list of cities and facilities are required, and must be developed as two text files, cityindix.txt and CostUniF.txt. The files must be presented in the format shown in the appendices. These two data files must be stored under the PCEM folder in the computer's C Drive.

e. The city index file has Houston as the base city, with value of 1.00. All the other city indexes in the file have values based on their costs deviation from Houston.

f. The escalation factor has to be determined separately from this estimate generation process.

g. The computer program is developed for a PC with Microsoft® Windows® 2000 operating systems.

6.4 PCEM PROGRAM EXECUTION

Figure 6.2 illustrates the five steps involved with the PCEM program execution.

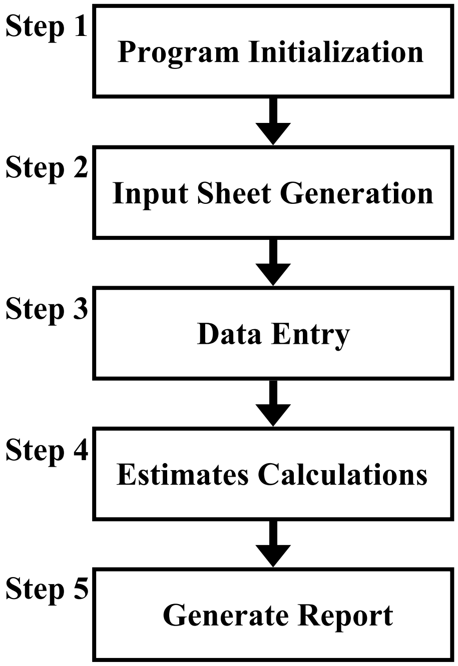
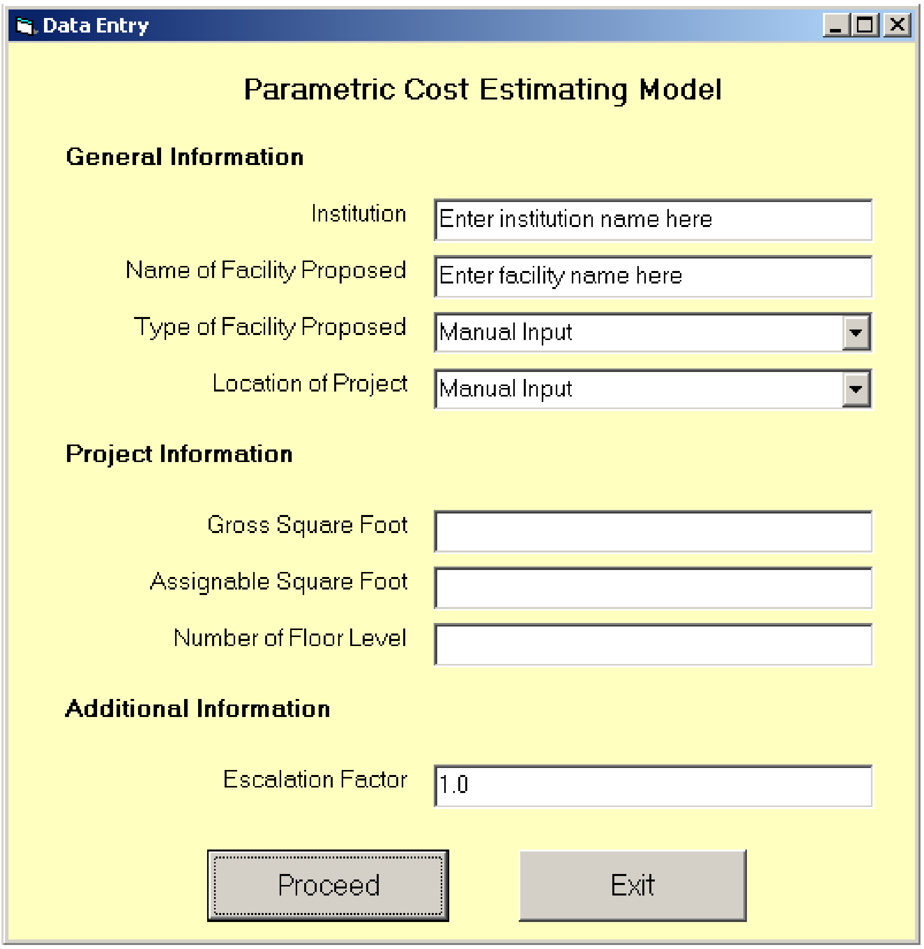


Figure 6.2. The Five Steps Involved with the PCEM Program Execution.

The following subsections discuss the detailed processes in each step.

6.4.1 PCEM Program Initialization and Input Sheet Generation

The estimating process starts with the program initialization, which launches the PCEM computer program. The program requires two data files, cityindex.txt and CostUniF.txt. The program accesses the two data files and their contents are used to develop the lists of cities and facility types for the user to select from. The city indexes and UniFormat ratios are also extracted for the subsequent use in the cost adjustments and calculations. With the two files read, the data input sheet or the Data Entry Window is generated by the program. Figure 6.3 shows this Data Entry Window.



Data Entry

Parametric Cost Estimating Model

General Information

Institution: Enter institution name here
Name of Facility Proposed: Enter facility name here
Type of Facility Proposed: Manual Input
Location of Project: Manual Input

Project Information

Gross Square Foot:
Assignable Square Foot:
Number of Floor Level:

Additional Information

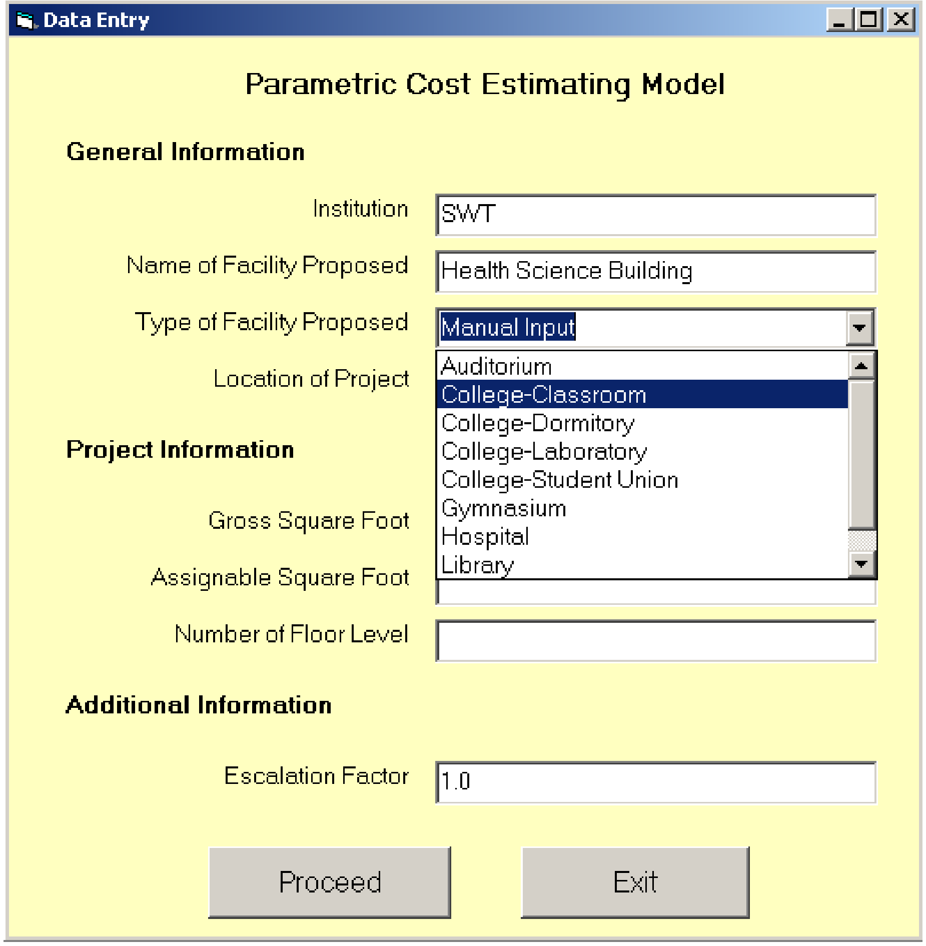
Escalation Factor: 1.0

Figure 6.3. PCEM Data Entry Window.

6.4.2 PCEM Data Entry

With the Data Entry Window generated, the user can then proceed to enter the required data into the appropriate fields on the form.

The information required on the data entry form can be divided into three main types: general information, project information, and additional information. Under the general information, the institution name is required along with the facility name. In these two fields, the names are typed in by the user. The next two fields provide drop down lists for the facility types and cities. These two drop down lists can be seen in Figures 6.4 and 6.5.



Parametric Cost Estimating Model

General Information

Institution	SWT
Name of Facility Proposed	Health Science Building
Type of Facility Proposed	Manual Input
Location of Project	

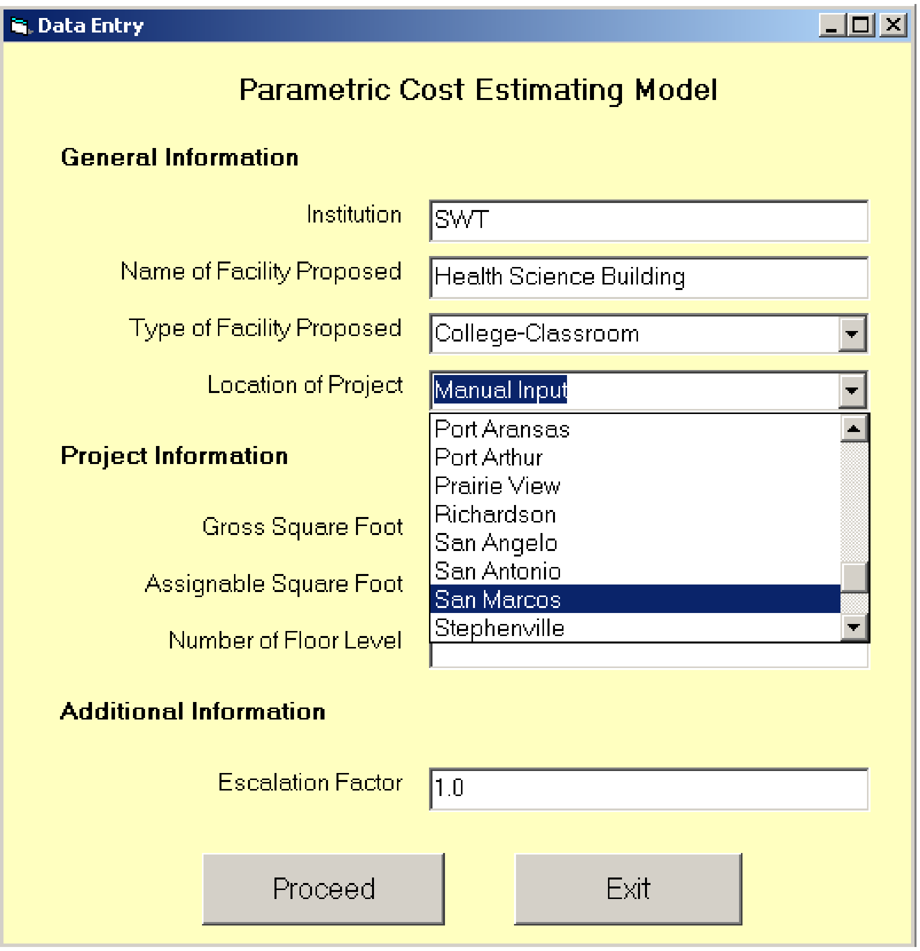
Project Information

Gross Square Foot	
Assignable Square Foot	
Number of Floor Level	

Additional Information

Escalation Factor	1.0

Figure 6.4. Facility Type Drop Down List



Data Entry

Parametric Cost Estimating Model

General Information

Institution: SWT

Name of Facility Proposed: Health Science Building

Type of Facility Proposed: College-Classroom

Location of Project: Manual Input

Project Information

Gross Square Foot

Assignable Square Foot

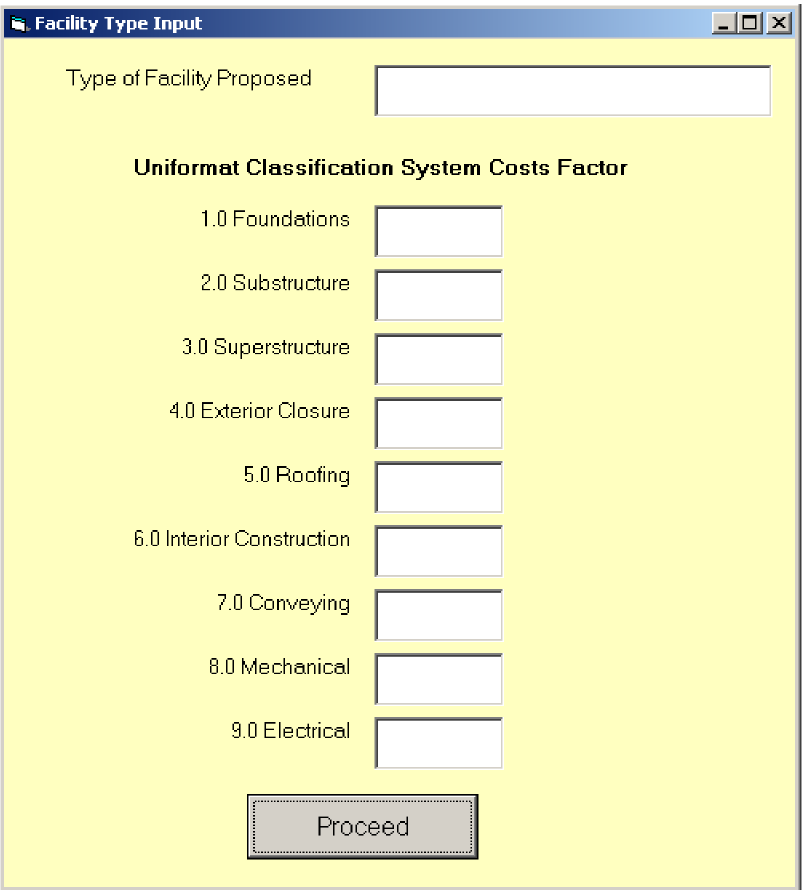
Number of Floor Level

Additional Information

Escalation Factor: 1.0

Figure 6.5. City Drop Down List

To complete the facility type field and the location field, the user can pick the facility type and location from the drop down lists. Alternatively, if none of the items from the list is selected, the program will display the appropriate form for inputting the new city data or facility type data. The two information request forms or windows are illustrated in Figures 6.6 and 6.7 for new city information and new facility type information, respectively.



Facility Type Input

Type of Facility Proposed

Uniformat Classification System Costs Factor

1.0 Foundations

2.0 Substructure

3.0 Superstructure

4.0 Exterior Closure

5.0 Roofing

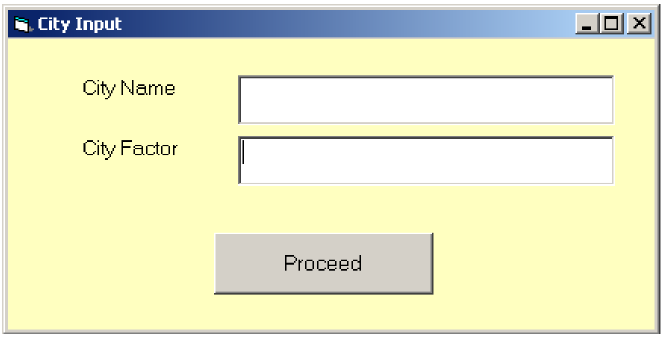
6.0 Interior Construction

7.0 Conveying

8.0 Mechanical

9.0 Electrical

Figure 6.6. New Facility Type Information Request Window



City Input

City Name

City Factor

Figure 6.7. New City Information Request Window

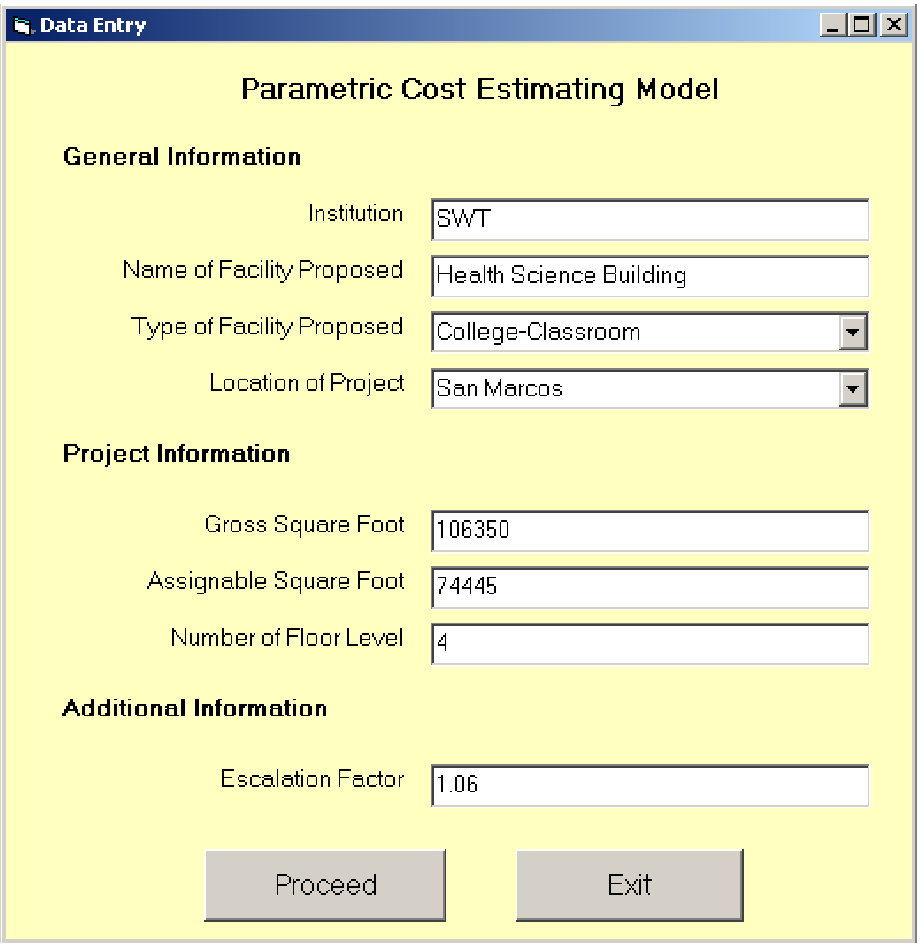
In the new facility type information request window or form, the user enters the new facility type name as defined by the user and the associated cost breakdown in ratio for each of the cost category listed. For example, if the foundations cost for the facility is estimated to be 10% of the total building construction cost, the value of 0.10 is entered into the field for Item 1.0 Foundations. The rest of the form is completed in this manner.

In the new city information request window or form, the new city name is entered along with the appropriate city factor. Since the base city for the model is Houston, with value of 1.00. If the construction cost associated with the new city is expected to be 10% less than Houston, a value of 0.90 is entered into the city factor field indicating that the predicted cost should be adjusted to 90% of the predicted value.

The Project Information section requires the Gross Square Footage, Assignable Square Footage, and the Number of Floor Level information of the proposed facility. The Additional Information required is the anticipated escalation factor for adjusting the estimated to the future time period. The program assigns a default value of 1.0, indicating no cost escalation is considered in the estimate, and the estimate produced is in 2001 dollars. This escalation factor can be determined by dividing the estimated index when construction is anticipated to begin with the index of the base year, which is 2001. For example, if the construction is planned for year 2003, and the index is forecasted to be

132.5, while the 2001 index is at 125.1, therefore the escalation factor can be calculated as 1.06 (132.5 divided by 125.1).

Once all the data are entered, the user clicks on the button "Proceed" to execute the estimate calculations. An "Exit" button is also provided to quit the program. Figure 6.8 shows a sample completed Data Entry Window.



Data Entry

Parametric Cost Estimating Model

General Information

Institution: SWT
Name of Facility Proposed: Health Science Building
Type of Facility Proposed: College-Classroom
Location of Project: San Marcos

Project Information

Gross Square Foot: 106350
Assignable Square Foot: 74445
Number of Floor Level: 4

Additional Information

Escalation Factor: 1.06

Figure 6.8. A Sample Completed PCEM Data Entry Window.

6.4.3 PCEM Estimate Calculations

The following calculations are involved in the estimate computation.

First, the computer program calculates the building cost in dollars per gross square foot. This calculation is based on the parametric cost estimating equation derived in Chapter 5, which is also displayed below.

$$\text{Cost estimate in \$ per GSF}$$
$$= 202.245 + 15.740 \ln(\text{number of floor}) - 126.196\ (\text{usage ratio}) \quad (\text{EQ } 6.1)$$

The building parameter values required for using the equation are the number of floor and the usage ratio. The usage ratio is computed by dividing the gross square footage by the assignable square footage. However, the estimated $perGSF calculated by the equation is in base dollar of Houston 2001, and therefore must be adjusted to the new location by the City Index and the Escalation Factor. The escalation factor has been manually entered, while the city index is retrieved for the data file based on the city selected, or if the values are manually enter, the newly entered index is used. The adjustment equation is displayed below.

$$\text{Adjusted \$perGSF} = \text{\$perGSF} \times \text{City Index} \times \text{Escalation Factor} \quad (\text{EQ } 6.2)$$

In addition, the metric value of this unit cost information is also provided, in dollars per square meter. This values is determined by dividing the \$perGSF value with 0.093 (1 SF = 0.093 m^2).

$$\mathit{\$perGSM} = \mathit{Adjusted\ \$perGSF} \div 0.093 \qquad \text{(EQ 6.3)}$$

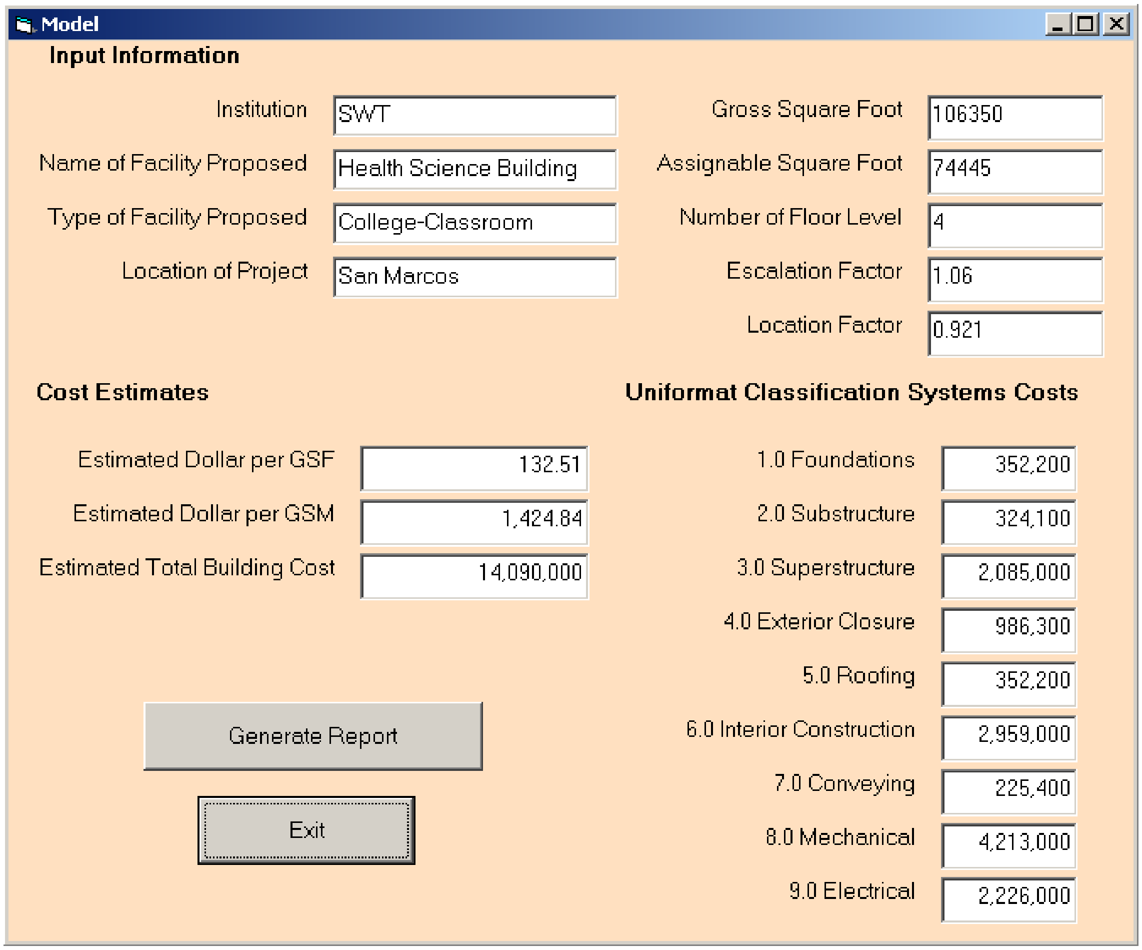
Next the total cost of the building can be computed by multiplying the \$perGSF by the total gross square footage of the building.

$$\mathit{Total\ Building\ Construction\ Cost} = \mathit{Adjusted\ \$perGSF} \times \mathit{GSF} \qquad \text{(EQ 6.4)}$$

The last steps are to breakdown the total building costs into the nine UniFormat Cost Classification Systems. The ratios for the cost breakdown have been automatically retrieved based on the type of facility selected, or if new facility type information are entered, the newly entered values. To calculate the cost of each category, the total cost is multiplied by the corresponding ratio.

$$\begin{gathered}\mathit{UniFormat\ Costs\ System\ (i)} = \mathit{Total\ Building\ Construction\ Cost} \times \mathit{ratio(i)} \\ \mathit{where\ i\ ranges\ from\ 1\ to\ 9\ for\ the\ 9\ systems} \qquad \text{(EQ 6.5)}\end{gathered}$$

Figure 6.9 illustrates a sample output sheet, or Model Window, generated by the PCEM program. All the computed values discussed are presented on this Model Window. In addition, a “Generate Report” button is also provided to generate this estimates information as a text file. By clicking on this button, the model generates a text file named EstimateReport.txt, which is placed in the PCEM folder in the C Drive. A message prompt, as shown in Figure 6.10, will also be presented to inform the user that the procedure is successfully accomplished. This text file output allows the user to print the output and subsequently save and stored this estimate as an electronic file for future database development. A sample of the text output is also presented in Figure 6.11.



Input Information

Institution	SWT
Name of Facility Proposed	Health Science Building
Type of Facility Proposed	College-Classroom
Location of Project	San Marcos
Gross Square Foot	106350
Assignable Square Foot	74445
Number of Floor Level	4
Escalation Factor	1.06
Location Factor	0.921

Cost Estimates

Estimated Dollar per GSF	132.51
Estimated Dollar per GSM	1,424.84
Estimated Total Building Cost	14,090,000

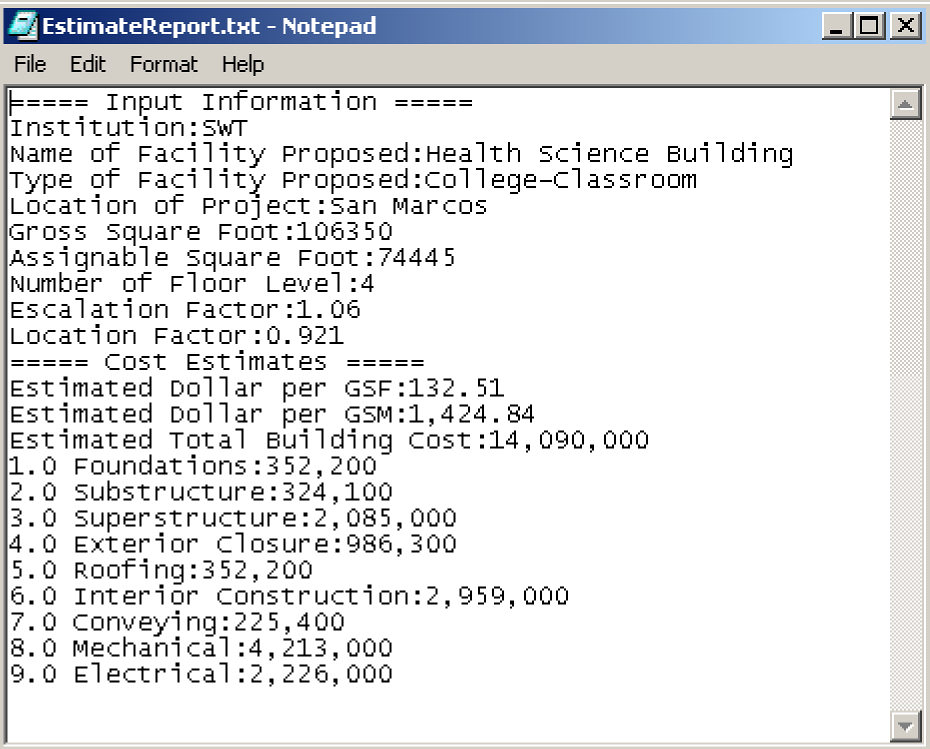
Uniformat Classification Systems Costs

1.0 Foundations	352,200
2.0 Substructure	324,100
3.0 Superstructure	2,085,000
4.0 Exterior Closure	986,300
5.0 Roofing	352,200
6.0 Interior Construction	2,959,000
7.0 Conveying	225,400
8.0 Mechanical	4,213,000
9.0 Electrical	2,226,000

Figure 6.9. Sample PCEM Model Window.

Estimate report was generated and placed in C:\PCEM

Figure 6.10. Report Generated Message Prompt



```
===== Input Information =====
Institution:SWT
Name of Facility Proposed:Health Science Building
Type of Facility Proposed:College-Classroom
Location of Project:San Marcos
Gross Square Foot:106350
Assignable Square Foot:74445
Number of Floor Level:4
Escalation Factor:1.06
Location Factor:0.921
===== Cost Estimates =====
Estimated Dollar per GSF:132.51
Estimated Dollar per GSM:1,424.84
Estimated Total Building Cost:14,090,000
1.0 Foundations:352,200
2.0 Substructure:324,100
3.0 Superstructure:2,085,000
4.0 Exterior Closure:986,300
5.0 Roofing:352,200
6.0 Interior Construction:2,959,000
7.0 Conveying:225,400
8.0 Mechanical:4,213,000
9.0 Electrical:2,226,000
```

Figure 6.11. Sample Text File Output

6.5 CHAPTER SUMMARY

This chapter presents the development of a computer program as a tool for conceptual cost estimating of building construction cost. The program is built on the parametric cost estimating model developed in Chapter 5. In addition to the program, two data files are also needed to run the program. The two data files are the cityindex.txt and CostUniF.txt. These files contain the city information and facility type information that is to be used for model execution and estimate computation.

The PCEM program is developed with Microsoft® Visual Basic® Programming System Version 6.0., as a stand alone application with GUI to facilitate usage. Many benefits can be realized from this computerized approach, among them are the quick, efficient and computational error free estimate generation. The PCEM program represents the computerized approached to conceptual cost estimating, it addresses the needs for standardization and systematized approach to conceptual cost estimating.

Chapter 7, presents the final chapter of the dissertations, which discusses the research conclusions and recommendations.

CHAPTER 7
CONCLUSIONS AND RECOMMENDATIONS

7.1 REVIEW OF RESEARCH OBJECTIVES

As presented in Chapter 1, the main objective of this research is to develop an accurate and practical method of systematic conceptual cost estimating that can be used by organizations involved in the planning and execution of building construction projects.

Three sub-objectives are also identified as follows:

1. To develop a parametric cost estimating model for conceptual cost estimating of building construction projects.

2. To identify and assess the relative importance of the significant building characteristics or parameters to be incorporated into a cost estimating model to improve the model's cost estimating performance in the early phase of the project development.

3. To develop a computer program, based on the developed parametric cost estimating model, as a tool for performing the conceptual cost estimating.

7.2 RESEARCH CONCLUSIONS

The main conclusions of this study are:

1. Development of Parametric Model for Conceptual Cost Estimating

Literature reviews for this research represents a very comprehensive discussion on various aspects of the parametric method. The application of the parametric method to building construction conceptual cost estimating is demonstrated through the development of the parametric cost estimating model. The model development involved data collection, database development, data preparation, and data analyses. These processes are performed in this study and have lead to the successful development of a parametric cost estimating model. From the model development discussions and validation results, it may be concluded that the developed model performed satisfactorily. For the model development data set, the percent cost deviations are quite low. The proportion of the projects having less then 20% cost deviations is 72%, and only 3.5% of the project cases have more then 50% cost deviations. For the validation data set, the results of the prediction for all the cases are within a 15% cost deviations. In addition, the model is statistically significant at the .05 level (F = 19.624, df = 2, 111) with $R^2 = 0.261$ and all the regression assumptions have been satisfied. Therefore, the model can be considered a reliable tool for performing the conceptual cost estimating.

2. Identification of Significant Building Parameters

Three building parameters were initially selected as the independent variables for predicting the single dependent variable, i.e. building construction cost in dollar per gross square foot. The three parameters are the gross square footage, number of floors, and the building usage ratio. The gross square footage and the number of floors are the parameters related to the scale of the building. The usage ratio, which is the ratio between the building assignable square footage and gross square footage, is proposed as a quality related measure variable for building construction projects. To enhance the logic of the developed model and to improve the analysis, the variables of gross square footage and number of floors were transformed. Subsequent multiple regression analysis concludes that the only significant variables for the parametric cost estimating model are the number of floors and the building usage ratio. The building usage ratio is also identified as the most powerful predictor for estimating the building unit cost.

3. Conceptual Cost Estimating Tool Development

To facilitate the implementation and usage of the parametric method, a computer program is envisioned and developed based on the parametric cost estimating model. The computer program is successfully developed using Microsoft® Visual Basic® Programming System with graphical user interface

(GUI) to facilitate the running of the applications. This computer program demonstrates a computerized approach to conceptual cost estimating. It facilitates the understanding of the concepts and implementation of the parametric cost estimating method to conceptual cost estimating of building construction cost.

Based on the stated research objectives and research findings, it may be concluded that this research has successfully accomplished all its objectives.

7.3 RECOMMENDATIONS FOR FUTURE RESEARCH

Several recommendations are offered concerning future research:

1. There is a need to identify additional building parameters, such as framing types and typical floor-to-floor heights, that may influence building construction costs, and incorporating them into the model so as to enhance the model's cost estimating performance.
2. To enhance the overall performance of the project conceptual cost estimating effort, other cost categories of a building construction project, such as architectural and engineering fees and contingencies, should also be investigated. Parametric cost estimating models can be similarly developed to estimate their costs.

3. There is a need to develop a list of standard institutional building types that have significant cost differences. The facilities may be classified based on the type of the major portion of the assignable area or the composition of the assignable areas. In this way, the analysis can accommodate mixed use facilities. The systematic classification of a facility into the standard building type should be based on some quantitative measure to minimize the subjectivity in the classification process and possibility of classification errors.
4. There is a need to develop proprietary indexes for use by THECB and other institutions in adjusting the cost for different locations and cost escalation, in addition to using those indexes from published sources.
5. Currently, THECB cost data are based on construction application forms submitted for project approval. There should be a research effort to perform additional data collection, to collect the as-built project information and the actual costs incurred for the projects.
6. Additional research may be directed to other building types not addressed in this research, such as parking garage buildings and apartment type buildings.
7. This research methodology should be applied to similar organizations in other states and to other construction related public organizations to determine the applicability of the

methodology developed and to confirm or assess any differences in the research findings.

7.4 RESEARCH CONTRIBUTIONS

This research investigation was exploratory in nature and contributes to the body of knowledge by expanding previous research investigations on parametric cost estimating for building construction projects. The focus of this research has been to identify the building parameters that significantly influence the cost of building construction and the development of the parametric cost estimating model, based on the parameters identified to demonstrate the successful application of the parametric method in conceptual cost estimating of building construction projects.

The first major contribution for this research is the identification of new building parameter, space usage ratio, as a significant cost driver for building construction cost. The identification of this new significant building parameter indicates the importance of this parameter. The inclusion of space usage ratio in the parametric cost estimating model developed in this study is shown to significantly improve the cost estimating model performance. This finding contributes to the current body of knowledge on conceptual cost estimating, and may serve as a useful guide for future data collection effort and cost estimating model development.

The second major contribution is the successful development of the parametric cost estimating model based on the parameters identified in the data analysis. The successful model development not only demonstrates the potential application of the parametric cost estimating method to conceptual cost estimating of building construction projects, but also has lead to the development of a tool, the PCEM computer program, that can be implemented to perform building conceptual cost estimates. The PCEM can serve as a valuable conceptual cost estimating tool for THECB and all other higher education institutions in Texas, that currently have no such tools.

The third contribution is the development of the electronic database for THECB for all their new building construction projects built in Texas from 1990-2000. This database serves as the preliminary effort at database development for THECB. It has demonstrated to be very useful for this research in developing the parametric cost estimating model for estimating the building construction cost per gross square footage. It is deemed that the database can be further analyzed to investigate other patterns in the data or cost trends. With this initial development paving the way for future data collection, it is anticipated that more data will be collected that will allow for an even better analysis than current procedures. It should be noted that the availability of data has been and still is one of the major problems hindering most research efforts.

Appendix A Historical Project Data

This Appendix contains the 168 new building construction project data collected from Texas Higher Education Coordinating Board (THECB) in 2001.

The Coordinating Board office is located at 1200 East Anderson Lane, Austin, TX 78752. The mailing address is P.O. Box 12788, Austin, TX 78711. The Coordinating Board offices can be contacted by phone at (512) 427-6101 or (512) 427-6127 (fax).

ID	stituition	Project Name	Year	Location	Facility Type	Floor	GSF	ASF	Ratio
2	SWT	Health Science Building	1990	San Marcos	Classroom & Laboratory	4	106350	74445	0.7
3	A&M U	Research Laboratory-Support Building	1990	College Station	Research Laboratory	1	4900	3920	0.8
4	UH Main	Athletic Facility	1990	Houston	Athletic Facility	3	58000	45000	0.776
5	UT Pan American	Academic Services Building	1990	Edinburg	Classroom & Laboratory	2	61075	42498	0.696
6	UT Pan American	Allied Health Annex	1990	Edinburg	Classroom & Laboratory	2	23975	15584	0.65
7	TSTC Harlingen	Engineering Graphics Technology Building	1990	Harlingen	Classroom & Laboratory	1	30844	24500	0.794
9	UT Dallas	Cecil and Ida Green Center	1991	Dallas	Administrative/Office	2	18688	10469	0.56
10	Prairie View A&M	Cooperative Extension Building	1991	Prairie View	Administrative/Office	1	21391	15225	0.712
11	SWT	Recreational Sports Building	1991	San Marcos	Athletic Facility	1	81272	62285	0.766
12	A&M U	Companion Animal Geriatric Center	1991	College Station	Administrative/Office	1	5962	3875	0.65
13	TSTC Sweetwater	Applied Technology Education Center Building	1991	Sweetwater	Classroom & Laboratory	1	20504	15065	0.735
14	TSTC Waco	Aerospace Technology Building	1991	Waco	Classroom & Laboratory	1	35000	29750	0.85
15	UT Arlington	Science Building Phase I Expansion	1991	Arlington	Classroom & Laboratory	3	62075	37191	0.599
16	A&M U	Field Lab, Animal Facility and Storage Facility	1992	Overton	Laboratory & Storage	1	9000	7500	0.833
17	UT Austin	Balcones Research Center-Warehouse Building	1991	Austin	Library	.	27500	24390	0.887
18	UT Austin	Parking Garage No. 2	1992	Austin	Parking Garage	5	213370	.	.
19	A&M U	Texas Beef Industry Center	1992	College Station	Research Laboratory	1	27085	24660	0.91
20	TSTC Amarillo	Aerospace/Industrial Technologies Center	1992	Amarillo	Classroom & Laboratory	1	37040	31100	0.84
21	A&M U	Regional Vocational Training Center	1992	San Antonio	Classroom & Laboratory	1	15000	13400	0.893
22	UT San Antonio	Transition Building	1992	San Antonio	Administrative/Office	1	11520	8640	0.75
23	TSTC Harlingen	Allied Health Technology Building	1992	Harlingen	Classroom & Laboratory	1	33743	25371	0.752
24	TSTC Harlingen	Aviation Technology Building	1992	Harlingen	Classroom & Laboratory	1	45269	34037	0.752
25	TSTC Harlingen	Autobody/Automotive Technology Building	1992	Harlingen	Classroom & Laboratory	1	36660	27564	0.752
26	Tarleton	Student Development Center	1992	Stephenville	Student Center	3	100760	67625	0.671
27	TSTC Harlingen	Student Housing	1992	Harlingen	Apartment Buildings	2	39424	20992	0.532
28	TX AES	Mariculture Lab Facility	1992	Port Aransas	Research Laboratory	1	13100	10850	0.828
29	TSTC Waco	Student Recreation Complex	1992	Waco	Athletic Facility	1	28000	25000	0.893
30	A&M Galveston	Physical Education Facility	1993	Galveston	Athletic Facility	1	28400	21445	0.755
31	A&M Corpus Christi	Classroom/Laboratory Building	1993	Corpus Christi	Classroom & Laboratory	4	78208	45283	0.579
33	TSTC Amarillo	Activity Center	1993	Amarillo	Student Center	1	26240	20624	0.786
35	UH Main	Athletic/Alumni	1993	Houston	Athletic Facility	4	222000	187288	0.844
36	UT Austin	Parking Garage at West 7th & Lavaca St.	1994	Austin	Parking Garage	2	52550	49364	0.939
37	Lamar-Port Arthur	Multipurpose Center	1994	Port Arthur	Student Center	4	31530	22486	0.713
38	UH Downtown	Student Life Building	1994	Houston	Athletic Facility	3	33600	29180	0.868
39	Angelo SU	State Disaster Recovery and Operations Data Center	1994	San Angelo	Classroom & Laboratory	3	85000	56950	0.67
40	UT San Antonio	Academic Building	1994	San Antonio	Classroom & Office	4	204790	122220	0.597
41	UT M Galveston	Lee Hage Jamail Student Center	1994	Galveston	Student Center	3	18306	11718	0.64
42	UT San Antonio	University Center Expansion	1994	San Antonio	Student Center	2	98840	60280	0.61
43	A&M U	Special Events Center	1995	College Station	Multi-Purpose	2	230000	177821	0.773
44	SWT	Student Center and Bookstore	1995	San Marcos	Student Center	5	201001	151957	0.756
45	UT El Paso	Classroom and Faculty Office Building	1995	El Paso	Classroom & Office	3	129840	79011	0.609
46	UT Pan America	Engineering Building	1995	Edinburg	Classroom & Laboratory	3	121184	73742	0.609

ID	Building	Fixed Equipment	Site Development	Total Construction	Movable Equipment	Fees	Contingency	Administrative	Total	$/GSF
2	$10,103,300			$10,103,300	$1,010,300	$732,000	$354,000	$151,500	$12,351,100	$95.00
3	$230,000	$20,000	$690,000	$850,000	$20,000	$51,000	$15,000	$64,000	$1,000,000	$46.94
4	$3,900,000	$100,000	$500,000	$4,500,000	$637,000	$490,000	$580,000	$293,000	$6,500,000	$67.24
5	$4,525,625	$306,375	$53,000	$4,885,000	$1,250,000	$313,000	$200,000	$448,200	$7,096,200	$74.10
6	$1,817,000	$137,600	$19,800	$1,974,400	$500,000	$130,700	$79,500	$165,800	$2,850,400	$75.79
7	$1,200,000			$1,200,000	$78,800	$61,200	$50,000	$10,000	$1,400,000	$38.91
9	$2,040,000			$2,040,000	$0	$135,300	$117,600	$140,100	$2,433,000	$109.16
10	$1,588,228	$78,748	$223,024	$1,890,000	$362,000	$115,000	$41,625	$157,143	$2,565,768	$74.25
11	$6,139,800			$6,139,800	$326,414	$443,786	$200,000	$90,000	$7,200,000	$75.55
12	$341,000	$9,000	$38,000	$388,000	$34,390	$23,500	$8,500	$45,610	$500,000	$57.20
13	$562,777	$305,728	$54,902	$923,407	$0	$54,424	$97,012	$25,157	$1,100,000	$27.45
14	$1,598,200		$105,000	$1,694,220	$0	$116,380	$170,000	$19,390	$1,999,990	$45.66
15	$7,200,000	$1,922,000	$808,000	$9,930,000	$750,000	$853,000	$356,000	$571,000	$12,460,000	$115.99
16	$400,000	$50,000	$21,000	$471,000	$90,000	$28,260	$11,000	$49,740	$650,000	$44.44
17	$1,529,000			$1,529,000	$0	$0	$61,000	$10,000	$1,600,000	$55.60
18	$4,043,000	$10,000	$50,000	$4,103,000	$0	$233,000	$275,000	$122,606	$4,733,606	$18.95
19	$953,798	$97,500	$201,900	$1,253,198	$80,000	$77,752	$27,882	$61,168	$1,500,000	$35.22
20	$1,562,000	$16,000	$100,000	$1,678,000	$0	$107,000	$200,000	$15,000	$2,000,000	$42.17
21	$826,000	$10,000	$30,000	$866,000	$0	$86,000	$20,000	$28,000	$1,000,000	$55.07
22	$338,000		$62,000	$400,000	$30,000	$28,000	$10,000	$32,000	$500,000	$29.34
23	$2,193,295			$2,193,295	$0	$132,000	$172,705	$2,000	$2,500,000	$65.00
24	$1,358,070			$1,358,070	$914,430	$89,500	$136,000	$2,000	$2,500,000	$30.00
25	$733,200	$100,000	$40,000	$873,200	$36,300	$50,000	$40,000	$500	$1,000,000	$20.00
26	$9,728,000	$198,000	$1,234,000	$10,710,000	$2,495,000	$642,600	$241,000	$675,154	$14,763,754	$96.55
27	$1,083,696			$1,083,696	$80,000	$74,500	$61,804	$0	$1,300,000	$27.49
28	$1,050,000	$716,000	$160,000	$1,926,000	$50,000	$134,200	$43,300	$131,400	$2,284,900	$80.15
29	$1,623,400		$115,000	$1,788,400	$0	$114,300	$0	$0	$1,902,700	$57.98
30	$2,665,000	$53,000	$520,000	$3,238,000	$0	$198,000	$46,000	$102,000	$3,584,000	$93.84
31	$7,843,358	$580,350	$526,292	$8,950,000	$664,000	$554,500	$202,000	$614,500	$10,985,000	$100.29
33	$1,140,600	$64,300	$38,850	$1,243,750	$0	$235,600	$69,550	$11,600	$1,560,500	$43.47
35	$20,103,000		$2,840,000	$22,943,000	$2,482,000	$1,932,000	$500,000	$1,243,000	$29,100,000	$90.55
36	$928,590			$928,590	$0	$68,000	$40,692	$41,000	$1,078,282	$17.67
37	$2,779,470	$50,000	$69,800	$2,899,270	$75,000	$159,850	$30,000	$1,500	$3,165,620	$88.15
38	$2,650,000		$100,000	$2,750,000	$199,700	$195,700	$50,000	$165,600	$3,361,000	$78.87
39	$6,375,000	$2,523,000		$8,898,000	$595,000	$964,510	$918,490	$120,000	$11,496,000	$75.00
40	$18,300,000	$720,000	$580,000	$19,600,000	$4,355,000	$1,100,000	$915,000	$1,030,000	$27,000,000	$89.36
41	$1,763,140	$167,860	$175,000	$2,106,000	$227,000	$140,000	$186,000	$191,000	$2,850,000	$96.31
42	$8,510,000	$210,000	$370,000	$9,090,000	$1,500,000	$560,000	$435,000	$415,000	$12,000,000	$86.10
43	$21,315,000	$2,555,000	$4,563,000	$28,433,000	$932,000	$1,510,000	$1,075,000	$1,475,000	$33,425,000	$92.67
44	$18,841,000			$18,841,000	$859,197	$1,249,952	$1,051,000	$324,851	$22,326,000	$93.74
45	$10,971,000	$700,000	$200,000	$11,871,000	$1,100,000	$850,000	$243,000	$381,000	$14,445,000	$84.50
46	$13,820,000	$180,000	$754,000	$14,754,000	$5,400,000	$1,000,000	$860,000	$1,040,000	$23,054,000	$114.04

ID	stituition	Project Name	Year	Location	Facility Type	Floor	GSF	ASF	Ratio
47	A&M U	Bush Presidential Library Center	1994	College Station	Classroom & Office	3	192229	119471	0.622
48	A&M Corpus Christi	Early Childhood Center	1995	Corpus Christi	Classroom & Office	2	53416	31378	0.587
49	A&M U	Dairy Products Teaching and Research Lab	1995	College Station	Research Laboratory	2	15282	10579	0.692
50	A&M Corpus Christi	University Services Center	1995	Corpus Christi	Administrative/Office	2	21284	14713	0.691
51	UNT HSC	Vivarium Expansion & Health Education Sciences Building	1995	Fort Worth	Health Care	1	80390	55354	0.689
52	TX AES	District Headquarters	1994	Fort Stockton	Administrative/Office	1	16425	11950	0.728
53	A&M Int U	Student Housing	1995	Laredo	Apartment Buildings	2	72000	48000	0.667
54	TxTech U	Special Collections Library	1995	Lubbock	Library	2	79800	59850	0.75
55	UT HSC San Antonio	Allied Health / Research Building	1995	San Antonio	Classroom & Laboratory	6	112608	67414	0.599
57	A&M Int U	Campus Facilities Phase 2	1995	Laredo	Classroom & Office	3	162394	99975	0.616
58	TxTech HSC	Health Sciences Center Library	1995	Lubbock	Library	4	88000	65000	0.739
59	Tx Woman's U	Movement Science Complex	1995	Denton	Athletic Facility	3	137681	93250	0.677
60	UT San Antonio	Downtown Campus Building	1995	San Antonio	Classroom & Office	4	112546	61354	0.545
61	UH Downtown	Academic/Student Services Building and Parking Garage	1995	Houston	Classroom & Office	5	177100	89000	0.503
62	TSTC Sweetwater	Recreational Sports & Health Facility	1995	Sweetwater	Athletic Facility	2	36058	31784	0.881
63	SWT	Parking Garage at Student Center/Bookstore	1995	San Marcos	Parking Garage	3	90000	.	.
65	TxTech U	Athletic Support Services Building	1995	Lubbock	Athletic Facility	1	4640	3572	0.77
66	UNT	Music & Fine Arts Education Building	1996	Denton	Performance Facility	2	72500	43500	0.6
67	UT San Antonio	Engineering/Biotechnology Building-Phase II	1995	San Antonio	Classroom & Laboratory	3	62120	38185	0.615
68	UT Pan American	Science Building	1996	Edinburg	Classroom & Laboratory	3	156701	87205	0.557
70	UT Austin	Parking Garage No.3	1996	Austin	Parking Garage	7	515510	494889	0.96
71	Tx Southern U	Science Building	1996	Houston	Classroom & Laboratory	4	145644	97919	0.672
72	Tx Southern U	Student Health Center	1994	Houston	Health Care	1	6800	5508	0.81
75	UT Austin	University Interscholastic League Building	1996	Austin	Administrative/Office	4	34604	22062	0.638
76	TSTC Harlingen	Science and Technology Building	1996	Harlingen	Classroom & Laboratory	2	60000	45114	0.752
77	TSTC Waco	Computer Applications Center	1996	Waco	Classroom & Laboratory	3	79500	59775	0.752
78	TSTC Waco	Fentress Center - Phase II (Automotive Technology)	1996	Waco	Classroom & Laboratory	1	40000	34000	0.85
79	UNT	Environmental Education, Science & Technology Building	1996	Denton	Classroom & Laboratory	3	108600	76020	0.7
80	Lamar-Port Arthur	Child Care & Development Center	1996	Port Arthur	Day Care Center	1	3360	2520	0.75
81	UT HSC Houston	Student Apartment	1996	Houston	Apartment Buildings	2	77624	70629	0.91
82	UT Dallas	Activity Center	1996	Dallas	Athletic Facility	2	86689	61326	0.707
83	Tx Southern U	School of Business	1996	Houston	Classroom & Office	3	78000	48700	0.624
84	UT San Antonio	Downtown Campus Building-Phase II	1997	San Antonio	Classroom & Office	4	145517	90871	0.624
86	UT SMC	North Campus - Phase III	1997	Dallas	Health Care	10	343851	138927	0.404
87	TxTech U	Residence Facilities-16 buildings	1997	Lubbock	Apartment Buildings	.	115466	92373	0.8
88	A&M Corpus Christi	University Center	1997	Corpus Christi	Student Center	3	95733	62116	0.649
89	TxTech U	C.A.Bassett Lab-Pulse Lab Addition	1997	Lubbock	Research Laboratory	.	8272	6650	0.804
90	Tx EES	Good Lab Practices Facility	1997	College Station	Research Laboratory	1	12092	8324	0.688
91	UT Pan American	International Trade and Technology Building	1997	Edinburg	Administrative/Office	1	21505	14581	0.678
93	A&M U	TTI Research Building	1997	College Station	Administrative/Office	3	60480	40320	0.667
94	A&M HSC	Education and Research Building	1997	Temple	Classroom & Laboratory	3	53569	31192	0.582
95	A&M U	Child Care Center	1997	College Station	Day Care Center	1	11288	7788	0.69

ID	Building	Fixed Equipment	Site Development	Total Construction	Movable Equipment	Fees	Contingency	Administrative	Total	$/GSF
47	$20,320,211	$582,548	$9,469,973	$30,372,732	$5,250,000	$1,903,345	$2,230,874	$1,987,699	$41,744,650	$105.71
48	$5,644,000		$354,000	$5,998,000	$543,000	$267,000	$135,000	$447,000	$7,390,000	$105.66
49	$1,901,000		$242,000	$2,143,000	$4,524	$149,500	$47,000	$155,976	$2,500,000	$124.39
50	$2,047,820		$150,000	$2,197,820	$361,990	$140,500	$48,700	$237,310	$2,986,320	$96.21
51	$6,964,418	$616,996	$390,258	$7,971,672	$786,121	$637,433	$517,298	$60,000	$9,972,524	$86.63
52	$711,000		$72,000	$788,000	$70,000	$49,000	$39,595	$19,900	$966,495	$43.29
53	$4,680,000	$210,000	$510,000	$5,400,000	$500,000	$342,000	$428,500	$419,500	$7,090,000	$65.00
54	$6,584,650	$776,100	$377,350	$7,738,100	$220,000	$492,000	$250,000	$99,900	$8,800,000	$82.51
55	$14,000,000	$460,000	$690,000	$15,150,000	$1,500,000	$930,000	$948,276	$471,724	$19,000,000	$124.33
57	$20,934,000	$760,000	$2,540,000	$24,234,000	$1,300,000	$1,132,800	$545,300	$1,626,900	$28,839,000	$128.91
58	$7,920,000	$480,000	$100,000	$8,500,000	$597,000	$528,000	$300,000	$75,000	$10,000,000	$90.00
59	$12,500,000	$625,000	$625,000	$13,750,000	$625,000	$962,500	$500,000	$0	$15,837,500	$90.79
60	$12,032,000	$320,000	$2,191,000	$14,543,000	$2,470,000	$939,000	$672,000	$721,000	$19,345,000	$106.91
61	$18,080,000		$165,000	$18,245,000	$866,000	$1,260,000	$655,000	$1,754,000	$22,780,000	$102.09
62	$1,524,000	$293,022	$91,440	$1,908,462	$123,309	$40,537	$152,400	$25,908	$2,250,616	$42.27
63	$2,500,000			$2,500,000	$0	$165,750	$50,000	$105,000	$2,820,750	$27.78
65	$452,500	$22,500	$40,000	$515,000	$0	$30,900	$24,600	$4,500	$575,000	$97.52
66	$10,585,000	$1,815,000	$550,000	$12,950,000	$0	$1,105,000	$250,000	$695,000	$15,000,000	$146.00
67	$8,231,000	$1,409,000	$310,000	$9,950,000	$2,710,000	$660,000	$660,000	$3,320,000	$17,300,000	$132.50
68	$17,550,000	$850,000	$1,200,000	$19,600,000	$2,600,000	$1,250,000	$700,000	$1,850,000	$26,000,000	$112.00
70	$9,080,000	$515,000		$9,595,000	$0	$572,000	$549,000	$284,000	$11,000,000	$17.61
71	$15,746,749	$1,372,932	$405,395	$17,525,076	$2,359,602	$2,512,556	$1,947,230	$179,468	$24,523,932	$108.12
72	$500,000	$50,000	$150,000	$700,000	$100,000	$70,000	$70,000	$20,000	$960,000	$73.53
75	$3,077,442		$249,558	$3,327,000	$125,000	$328,000	$312,000	$358,800	$4,450,800	$88.93
76	$3,650,000	$200,000		$3,850,000	$400,000	$250,000	$500,000	$0	$5,000,000	$60.83
77	$5,740,000	$140,000	$560,000	$6,440,000	$0	$350,000	$140,000	$70,000	$7,000,000	$72.20
78	$1,640,000	$40,000	$160,000	$1,840,000	$0	$100,000	$40,000	$20,000	$2,000,000	$41.00
79	$9,997,160	$1,730,500	$542,540	$12,270,200	$300,000	$935,060	$1,106,110	$282,000	$14,893,370	$92.05
80	$326,000	$3,500		$329,500	$0	$22,750	$12,000	$400	$364,650	$97.02
81	$2,496,600	$81,000	$335,900	$2,913,500	$0	$216,000	$402,551	$143,949	$3,676,000	$32.16
82	$8,270,000	$500,000	$400,000	$9,170,000	$134,000	$620,000	$993,000	$383,000	$11,300,000	$95.40
83	$9,251,730	$1,104,973	$408,214	$10,764,917	$1,380,000	$1,913,993	$710,218	$145,456	$14,914,584	$118.61
84	$19,500,000	$2,000,000	$100,000	$22,500,000	$6,500,000	$1,300,000	$700,000	$1,000,000	$32,000,000	$134.01
86	$53,861,483	$4,956,555	$3,349,580	$62,167,618	$4,641,000	$3,581,118	$7,600,151	$2,010,113	$80,000,000	$156.64
87	$9,500,000		$2,250,000	$11,750,000	$1,500,000	$840,500	$587,500	$322,000	$15,000,000	$82.28
88	$10,869,000	$17,000	$394,000	$11,280,000	$1,266,968	$700,000	$254,250	$987,782	$14,489,000	$113.53
89	$1,122,525		$85,000	$1,207,525	$0	$160,238	$104,205	$28,032	$1,500,000	$135.70
90	$800,000	$40,000	$86,000	$926,000	$0	$46,000	$22,000	$92,000	$1,086,000	$66.16
91	$2,075,000	$100,000	$425,000	$2,600,000	$0	$200,000	$200,000	$150,000	$3,150,000	$96.49
93	$6,652,800		$431,200	$7,084,000	$375,000	$455,000	$159,390	$426,610	$8,500,000	$110.00
94	$7,418,000	$950,000	$302,000	$8,670,000	$900,000	$512,000	$195,000	$723,000	$11,000,000	$138.48
95	$950,000		$10,000	$960,000	$0	$51,000	$18,000	$71,000	$1,100,000	$84.16

ID	stituition	Project Name	Year	Location	Facility Type	Floor	GSF	ASF	Ratio
96	A&M Texarkana	Aikin Building Expansion	1998	Texarkana	Classroom & Office	2	37100	24200	0.652
98	UT HSC San Antonio	Research Building in Texas Research Park	1997	San Antonio	Research Laboratory	4	94659	56747	0.599
99	UH Main	Music Building	1995	Houston	Performance Facility	3	141114	90723	0.643
101	UH Main	Communication Disorders/Psychology Building	1997	Houston	Administrative/Office	1	24500	15710	0.641
102	TxTech U	Commuter Parking Structure	1997	Lubbock	Parking Garage	2	200815	.	.
103	UNT	Speech & Hearing Clinic	1998	Denton	Administrative/Office	2	14800	10650	0.72
104	UT Dallas	Waterview Park Apartments-Phase VI	1998	Dallas	Apartment Buildings	3	87000	77760	0.894
105	UT Permian Basin	Visual Art Studios	1998	Odessa	Classroom & Laboratory	1	27254	19217	0.705
106	UH Main	Center for Public Broadcasting	1998	Houston	Administrative/Office	4	62300	41339	0.664
107	SWT	Art, Technology & Physics Complex	1999	San Marcos	Classroom & Laboratory	5	237591	157230	0.662
108	UT Tyler	Liberal Arts Complex	1995	Tyler	Performance Facility	3	124808	77224	0.619
109	UH Main	Charter School	1998	Houston	School	1	5489	4041	0.736
110	TxTech U	English, Philosophy & Education Complex	1999	Lubbock	Classroom & Office	3	202983	111640	0.55
111	UT Dallas	Callier Center for Communication Disorders Expansion	1998	Dallas	Administrative/Office	2	12114	7753	0.64
112	TxTech U	Athletic Academic Services Building	1999	Lubbock	Classroom & Office	.	14708	9560	0.65
114	TxTech HSC Midland	Physician Associate Program Building	1999	Midland	Classroom & Office	1	35000	22032	0.629
115	Tx Southern U	Student Recreational Center	1998	Houston	Athletic Facility	2	100300	78300	0.781
116	TX AES	Natural Resources Informatics Laboratory	1998	Temple	Administrative/Office	1	12200	8020	0.657
117	UT Austin	Parking Garage No. 4	1998	Austin	Parking Garage	5	211567	206166	0.974
118	A&M Prairie View	Science Building	1999	Prairie View	Classroom & Laboratory	4	158517	95133	0.6
119	Tarleton	Science Building	1999	Stephenville	Classroom & Laboratory	4	155188	92919	0.599
120	Tx EES	Turbomachinery Office Building	1999	College Station	Administrative/Office	2	9000	5845	0.649
122	TSTC Harlingen	Semiconductor Manufacturing Technology Building	1999	Harlingen	Classroom & Laboratory	1	28000	22000	0.786
123	UH Fort Bend	Academic Building	1999	Fort Bend	Classroom & Office	3	56851	35337	0.622
124	Sam Houston SU	Law Enforcement Management Institute (LEMIT)	1999	Huntsville	Classroom & Office	4	30850	16310	0.529
125	Sul Ross SU	Multi-Purpose Center	1999	Alpine	Multi-Purpose	2	81556	49775	0.61
126	A&M Int U	Student Development Center	1999	Laredo	Student Center	2	89900	64064	0.713
127	Angelo SU	Multi-purpose Center	1999	San Angelo	Multi-Purpose	2	132254	92547	0.7
128	SWT	Admissions Center Expansion	1999	San Marcos	Administrative/Office	1	5500	3850	0.7
129	Lamar-Beaumont	Technology Center	1999	Beaumont	Classroom & Office	2	22100	14040	0.635
130	UT Tyler	Longview Higher Education Center	1999	Longview	Classroom & Office	2	25060	15145	0.604
132	UT Permian Basin	Library/Lecture Center	1999	Odessa	Library	2	78827	57697	0.732
133	A&M Kingsville	Physical Conditioning Laboratory	1999	Kingsville	Athletic Facility	1	6000	5000	0.833
134	TxTech U	Athletic Training/Rehabilitation Center and Hall of Fame	1999	Lubbock	Classroom & Office	.	16760	11079	0.661
135	UT Dallas	Waterview Park Apartments Phase VII	1999	Dallas	Apartment Buildings	3	87000	77760	0.894
136	UT SMC	Radiation Oncology Center	1999	Dallas	Health Care	3	30000	18000	0.6
137	A&M Int U	Western Hemispheric Trade Center	1999	Laredo	Administrative/Office	2	55043	36293	0.659
138	UT MD Anderson	Faculty Center	1999	Houston	Administrative/Office	14	325000	232677	0.716
139	TSTC Harlingen	Field House expansion	1999	Harlingen	Athletic Facility	1	6000	5409	0.902
140	UT Pan American	Student Union	1999	Edinburg	Student Center	2	44722	29050	0.65
141	A&M Kingsville	Engineering Building	1999	Kingsville	Classroom & Laboratory	3	78439	47917	0.611
142	UT Arlington	Residence Hall	1999	Arlington	Dormitory	3	175300	122700	0.7

ID	g	Fixed Equipment	Site Development	Total Construction	Movable Equipment	Fees	Contingency	Administrative	Total	$/GSF
96	$3,626,000	$10,000	$185,000	$3,821,000	$628,000	$245,000	$85,000	$263,000	$5,042,000	$97.74
98	$12,840,000	$900,000	$1,500,000	$15,240,000	$650,000	$930,000	$1,325,000	$855,000	$19,000,000	$135.64
99	$18,495,549		$27,000	$18,522,549	$1,063,225	$1,810,910	$435,841	$1,284,675	$23,117,200	$131.07
101	$1,460,500	$78,250	$150,000	$1,688,750	$100,000	$67,000	$84,438	$116,888	$2,057,075	$59.61
102				$3,904,000	$0	$234,240	$0	$61,760	$4,200,000	$0.00
103	$1,354,500	$55,000	$10,000	$1,419,500	$0	$119,000	$140,900	$70,600	$1,750,000	$91.52
104	$3,655,365		$344,635	$4,000,000	$0	$0	$177,200	$122,800	$4,300,000	$42.02
105	$2,902,000	$56,000	$150,000	$3,108,000	$348,200	$208,100	$298,375	$187,325	$4,150,000	$106.48
106	$7,422,250	$979,000	$363,125	$8,764,375	$750,000	$1,198,725	$600,000	$564,900	$11,878,000	$119.14
107	$26,499,337		$2,800,000	$29,299,337	$2,192,221	$2,613,162	$2,173,613	$799,005	$37,077,338	$111.53
108	$14,310,000	$1,000,000	$1,100,000	$16,410,000	$1,400,000	$1,250,000	$360,000	$530,000	$19,950,000	$114.66
109	$386,600	$10,000	$70,000	$466,600	$10,000	$30,900	$25,902	$32,598	$566,000	$70.43
110	$28,823,586		$1,629,022	$30,452,608	$4,425,000	$3,109,976	$1,470,003	$1,434,913	$40,892,500	$142.00
111	$1,817,000	$375,000	$50,000	$2,242,000	$443,210	$188,090	$295,183	$151,000	$3,319,483	$149.99
112	$1,150,000		$190,000	$1,340,000	$165,000	$138,940	$119,000	$237,060	$2,000,000	$78.19
114	$3,500,000		$520,000	$4,020,000	$850,000	$460,500	$300,000	$369,500	$6,000,000	$100.00
115	$8,090,922	$322,000	$178,250	$8,591,172	$475,000	$1,312,588	$859,117	$50,000	$11,287,877	$80.67
116	$874,500		$180,000	$1,054,500	$0	$64,000	$23,700	$107,800	$1,250,000	$71.68
117	$5,331,191	$755,568	$865,331	$7,058,587	$20,000	$796,028	$727,859	$597,526	$9,200,000	$25.20
118	$19,676,934	$2,977,148	$865,918	$23,520,000	$1,100,000	$1,411,200	$530,000	$1,438,800	$28,000,000	$124.13
119	$17,320,000	$2,440,000	$3,650,000	$23,410,000	$1,100,000	$1,404,600	$527,000	$1,258,000	$27,699,600	$111.61
120	$880,000		$55,000	$935,000	$182,000	$65,100	$21,000	$125,900	$1,329,000	$97.78
122	$2,235,000	$479,000		$2,714,000	$0	$147,000	$137,900	$1,100	$3,000,000	$79.82
123	$6,570,000		$2,100,000	$8,670,000	$200,000	$701,800	$921,000	$507,200	$11,000,000	$115.57
124	$3,808,000	$285,000	$226,000	$4,319,000	$250,000	$321,580	$275,000	$420,000	$5,585,580	$123.44
125	$7,942,089	$607,400	$595,273	$9,144,762	$200,000	$617,271	$100,000	$60,000	$10,122,033	$97.38
126	$10,400,000	$1,000,000	$600,000	$12,000,000	$1,200,000	$720,000	$300,000	$780,000	$15,000,000	$115.68
127	$15,806,451	$1,649,900		$17,456,351	$264,268	$1,047,381	$300,000	$20,000	$19,088,000	$119.52
128	$455,000			$455,000	$38,697	$45,147	$19,951	$13,655	$572,450	$82.73
129	$2,268,650	$273,594	$101,700	$2,643,944	$0	$156,650	$72,300	$96,400	$2,969,294	$102.65
130	$2,527,000	$100,000	$621,000	$3,248,000	$375,000	$370,000	$352,700	$354,300	$4,700,000	$100.84
132	$10,329,420	$80,000	$1,388,100	$11,797,520	$1,315,240	$718,970	$1,463,170	$555,100	$15,850,000	$131.04
133	$450,000			$450,000	$64,500	$35,000	$10,000	$40,500	$600,000	$75.00
134	$1,435,000	$465,000	$548,000	$2,448,000	$545,000	$271,465	$240,000	$495,535	$4,000,000	$85.62
135	$3,655,365		$344,635	$4,000,000	$0	$0	$177,200	$122,800	$4,300,000	$42.02
136	$6,774,410	$800,000	$565,000	$8,139,410	$190,000	$846,000	$504,590	$320,000	$10,000,000	$225.81
137	$7,200,000	$160,000	$200,000	$7,560,000	$1,200,000	$453,600	$170,000	$616,400	$10,000,000	$130.81
138	$35,600,000		$356,000	$35,956,000	$3,500,000	$2,517,000	$2,120,000	$889,000	$44,982,000	$109.54
139	$500,000			$500,000	$25,000	$45,000	$30,000	$0	$600,000	$83.33
140	$5,000,000	$150,000	$100,000	$5,250,000	$550,000	$400,000	$448,038	$351,962	$7,000,000	$111.80
141	$10,673,000	$257,000	$870,000	$11,800,000	$1,500,000	$715,000	$265,000	$720,000	$15,000,000	$136.07
142	$15,975,000		$1,700,000	$17,675,000	$1,400,000	$900,640	$1,236,260	$788,100	$22,000,000	$91.13

ID	stituition	Project Name	Year	Location	Facility Type	Floor	GSF	ASF	Ratio
143	UT Brownsville	Life & Health Science Building	1999	Brownsville	Classroom & Laboratory	3	101378	60893	0.601
144	UT Austin	Psychology, Child Development & Family Relationships Building	1999	Austin	Classroom & Laboratory	4	180920	108552	0.6
145	UT Austin	New Student Housing	1999	Austin	Dormitory	6	304183	168177	0.553
146	A&M U	University Apartments Community Center	1999	College Station	Student Center	1	16022	10228	0.638
147	Tx EES	Coastal Engineering Laboratory Building	2000	College Station	Research Laboratory	1	21640	17892	0.827
148	A&M HSC	Regional Health Science Education Center in Temple	2000	Temple	Classroom & Office	4	62154	40093	0.645
149	UNT HSC	Parking Garage	1999	Fort Worth	Parking Garage	4	265750	.	.
150	A&M Corpus Christi	Science and Technology Building	2000	Corpus Christi	Classroom & Laboratory	4	67366	40131	0.596
151	UT Pan American	General Classroom/Computer Center Building	1999	Edinburg	Classroom & Office	3	108371	65470	0.604
152	UNT	University Gateway Center	2000	Denton	Classroom & Office	3	116000	77340	0.667
153	A&M Int U	Fine Arts Complex	2000	Laredo	Performance Facility	2	88250	52713	0.597
154	UT Austin	Applied Research Laboratory Expansion	1999	Austin	Research Laboratory	2	20222	13607	0.673
155	Sam Houston SU	General Classroom/Office Building	2000	Huntsville	Classroom & Office	4	61067	36715	0.601
156	Lamar-Orange	Library/Administration Building	1999	Orange	Administrative/Office	3	51465	38601	0.75
157	Tx Southern U	Student Health Center	1999	Houston	Health Care	1	8800	7200	0.818
158	UT Pan American	Student Housing	2000	Edinburg	Dormitory	2	59288	55498	0.936
159	SWT	Art/Technology/Physics Building	2000	San Marcos	Classroom & Laboratory	5	237591	157230	0.662
160	A&M U	West Campus Parking Garage	2000	College Station	Parking Garage	7	1022246	.	.
161	TxTech HSC Amarillo	Academic/Clinic Building	2000	Amarillo	Administrative/Office	5	157890	111456	0.706
162	UT HSC San Antonio	Medical Education Division-Harlingen of the RAHC	2000	Harlingen	Classroom & Office	3	86781	55298	0.637
163	UT HSC Houston	Brownsville Public Health Division-RAHC	2000	Brownsville	Classroom & Office	2	25000	15000	0.6
165	UT SMC Dallas	Student Housing	2000	Dallas	Apartment Buildings	3	159257	129420	0.813
166	TxTech HSC	Auditorium/Classroom	2000	Lubbock	Classroom & Assembly	4	59487	42830	0.72
167	UT Dallas	Student Housing-Phase VIII	2000	Richardson	Apartment Buildings	3	170000	125800	0.74
168	UT El Paso	Student Housing	2000	El Paso	Apartment Buildings	3	154000	123000	0.799
169	A&M Prairie View	Student Center	2000	Prairie View	Student Center	3	120000	85670	0.714
170	UT Dallas	Student Life Annex	2000	Richardson	Student Center	2	12100	9100	0.752
171	UT HSC San Antonio	Laredo Campus Extension	2000	Laredo	Classroom & Assembly	2	19995	11258	0.563
172	Tarleton	Student Housing	2000	Stephenville	Apartment Buildings	3	90654	68472	0.755
173	A&M Commerce	Student Housing	2000	Commerce	Apartment Buildings	2	100778	80300	0.797
174	UT SMC Dallas	Student Services Building	2000	Dallas	Athletic Facility	2	56000	44570	0.796
175	A&M U 'West TX	Events Center	2000	Canyon	Multi-Purpose	2	76599	49901	0.651
176	A&M U	Theriogenology (Equine Reproduction) Facility	2000	College Station	Research Laboratory	1	19468	11205	0.576
177	UT San Antonio	Child Care Center	2000	San Antonio	Day Care Center	1	11182	7976	0.713
178	UT El Paso	Larry K. Durham Sports Center	2000	El Paso	Athletic Facility	2	73000	52838	0.724
179	UT San Antonio	Recreation/Wellness Center	2000	San Antonio	Athletic Facility	2	69772	50416	0.723
180	UT Austin	Parking Garage South	2000	Austin	Parking Garage	5	369150	348461	0.944
181	UT Dallas	Engineering and Computer Science Complex	2000	Dallas	Classroom & Laboratory	4	140500	84315	0.6
182	TxTech U	Experimental Science Building Stage I	2000	Lubbock	Classroom & Laboratory	.	84140	47665	0.566
183	UT San Antonio	Downtown Campus Phase III	2000	San Antonio	Classroom & Laboratory	4	294938	177017	0.6
184	UT San Antonio	Academic Building III	2000	San Antonio	Classroom & Laboratory	3	190830	122324	0.641
185	UT MD Anderson	Basic Sciences Research Building Phase II	2000	Houston	Research Laboratory	19	485740	293315	0.604

ID	g	Fixed Equipment	Site Development	Total Construction	Movable Equipment	Fees	Contingency	Administrative	Total	$/GSF
143	$15,000,000	$500,000	$400,000	$15,900,000	$4,000,000	$897,900	$983,145	$718,955	$22,500,000	$147.96
144	$33,447,333	$1,490,000	$2,052,005	$36,989,338	$0	$2,899,359	$1,801,311	$1,878,272	$43,568,280	$184.87
145	$40,724,695	$660,772	$2,625,475	$44,010,942	$2,125,000	$2,460,000	$699,237	$2,435,725	$51,730,904	$133.88
146	$1,728,560		$255,992	$1,984,552	$152,625	$130,000	$46,000	$184,025	$2,497,202	$107.89
147	$2,960,900	$116,500	$290,000	$3,367,400	$100,000	$313,000	$75,735	$213,865	$4,070,000	$136.83
148	$6,950,000	$150,000	$700,000	$7,800,000	$385,000	$474,000	$176,000	$665,000	$9,500,000	$111.82
149	$7,973,000		$499,000	$8,472,000	$0	$560,500	$842,000	$0	$9,874,500	$30.00
150	$8,763,801	$111,407	$2,553,792	$11,429,000	$520,360	$730,700	$225,000	$823,940	$13,729,000	$130.09
151	$12,500,000	$300,000	$300,000	$13,100,000	$1,700,000	$880,000	$820,000	$700,000	$17,200,000	$115.34
152	$14,100,000	$200,000	$250,000	$14,550,000	$870,000	$1,050,000	$555,000	$395,000	$17,420,000	$121.55
153	$11,992,729	$2,067,909	$619,362	$14,680,000	$1,130,000	$954,200	$331,000	$904,800	$18,000,000	$135.89
154	$2,279,164	$11,800	$199,100	$2,490,064	$0	$215,424	$144,862	$240,050	$3,090,400	$112.71
155	$7,667,700			$7,667,700	$300,000	$536,739	$200,000	$22,000	$8,726,439	$125.56
156	$5,928,593	$5,000	$3,015,132	$8,948,725	$177,775	$726,500	$533,552	$213,448	$10,600,000	$115.20
157	$1,394,230	$100,000	$23,000	$1,517,230	$60,000	$160,190	$124,873	$57,241	$1,919,534	$158.44
158	$3,350,000		$650,000	$4,000,000	$110,000	$450,000	$275,000	$165,000	$5,000,000	$56.50
159	$31,349,428		$2,800,000	$34,149,428	$2,192,221	$4,544,052	$682,989	$799,005	$42,367,695	$131.95
160	$23,480,000	$1,428,000	$1,607,000	$26,515,000	$15,000	$2,750,000	$600,000	$120,000	$30,000,000	$22.97
161	$15,333,000	$600,000	$1,400,000	$17,333,000	$878,000	$2,455,000	$324,000	$835,000	$21,825,000	$97.11
162	$14,700,000	$500,000	$1,750,000	$16,950,000	$4,500,000	$2,173,327	$1,356,673	$20,000	$25,000,000	$169.39
163	$3,525,000			$3,525,000	$350,000	$745,836	$209,164	$170,000	$5,000,000	$141.00
165	$7,750,000	$50,000	$700,000	$8,500,000	$350,000	$752,500	$547,500	$350,000	$10,500,000	$48.66
166	$8,625,506	$418,350	$1,100,000	$10,143,856	$79,296	$1,562,904	$862,550	$556,777	$13,205,383	$145.00
167	$11,046,000	$50,000	$500,000	$11,596,000	$545,000	$950,000	$754,000	$155,000	$14,000,000	$64.98
168	$10,840,000	$70,000	$1,000,000	$11,910,000	$835,000	$1,030,000	$750,000	$475,000	$15,000,000	$70.39
169	$14,280,000	$340,000	$2,380,000	$17,000,000	$1,655,000	$2,483,500	$383,500	$478,000	$22,000,000	$119.00
170	$1,711,055	$50,000	$115,000	$1,876,055	$332,000	$378,525	$199,000	$30,000	$2,815,580	$141.41
171	$3,400,000	$10,092		$3,410,092	$50,000	$477,908	$60,000	$2,000	$4,000,000	$170.04
172	$3,491,600	$79,000	$799,400	$4,370,000	$970,000	$577,000	$98,000	$20,000	$6,035,000	$38.52
173	$4,038,000	$108,000	$848,000	$4,994,000	$1,125,000	$454,000	$112,000	$220,000	$6,905,000	$40.07
174	$7,770,000	$200,000	$460,000	$8,430,000	$840,000	$1,060,000	$540,000	$50,000	$10,920,000	$138.75
175	$8,877,516	$485,013	$1,434,424	$10,796,953	$0	$1,288,945	$887,947	$176,155	$13,150,000	$115.90
176	$1,175,000		$25,000	$1,200,000	$0	$120,000	$120,000	$60,000	$1,500,000	$60.36
177	$1,460,000	$76,000	$404,000	$1,940,000	$231,000	$417,000	$214,000	$98,000	$2,900,000	$130.57
178	$6,465,240	$100,000	$300,000	$6,865,240	$0	$829,752	$440,008	$35,000	$8,170,000	$88.56
179	$10,164,000	$137,000	$971,000	$11,272,000	$1,025,000	$1,665,000	$974,000	$439,000	$15,375,000	$145.67
180	$16,046,796	$418,108	$1,518,453	$17,983,357	$80,000	$2,459,654	$1,601,989	$375,000	$22,500,000	$43.47
181	$23,500,000	$500,000	$1,000,000	$25,000,000	$1,000,000	$2,804,000	$1,096,000	$100,000	$30,000,000	$167.26
182	$19,024,243	$2,148,757	$1,427,000	$22,600,000	$10,000,000	$4,455,820	$1,465,000	$1,479,180	$40,000,000	$226.10
183	$36,078,744	$1,670,000	$1,335,000	$39,083,744	$553,881	$1,600,801	$1,306,574	$455,000	$43,000,000	$122.33
184	$28,827,000	$657,000	$2,166,000	$31,650,000	$8,300,000	$5,262,241	$2,897,460	$890,299	$49,000,000	$151.06
185	$122,361,453			$122,361,453	$10,230,750	$16,835,159	$8,727,514	$6,445,124	$164,600,000	$251.91

Appendix B City Cost Indexes Data File

This Appendix contains the cost indexes data for various Texas cities. The city cost indexes have been adapted from *Means Building Construction Cost Data 2002*.This data file is accessed by the Parametric Cost Estimating Model (PCEM) Computer Program in the generation of the cities list and in adjusting the estimate produced to the respective city. The name assigned to this data file, which is a simple text file, is “cityindex.txt.”

```
Alpine:0.864
Amarillo:0.915
Arlington:0.933
Austin:0.921
Beaumont:0.945
Brownsville:0.875
Canyon:0.915
College Station:0.933
Commerce:0.887
Corpus Christi:0.892
Dallas:0.965
Denton:0.887
Edinburg:0.875
El Paso:0.884
Fort Bend:1.000
Fort Stockton:0.864
Fort Worth:0.933
Galveston:0.980
Harlingen:0.875
Houston:1.000
Huntsville:0.828
Kingsville:0.892
Laredo:0.875
Longview:0.835
Lubbock:0.905
Midland:0.890
Odessa:0.864
Orange:0.945
Overton:0.905
Port Aransas:0.892
Port Arthur:0.945
Prairie View:0.933
Richardson:0.965
San Angelo:0.855
San Antonio:0.935
San Marcos:0.921
Stephenville:0.904
Sweetwater:0.892
Temple:0.863
Texarkana:0.869
Tyler:0.905
Waco:0.904
```

Appendix C Facility Types and UniFormat Cost Breakdown Data File

This Appendix contains the list of facility types and the respective cost breakdown ratios by UniFormat Cost Classification System. The type of facility is first listed and the following numbers represent the ratios of the UniFormat Cost for the nine building system and the total building construction cost:

1.0 Foundations
2.0 Substructure
3.0 Superstructure
4.0 Exterior Closure
5.0 Roofing
6.0 Interior Construction
7.0 Conveying
8.0 Mechanical
9.0 Electrical

The ratios have been adapted from *Means Square Foot Cost 2002*. This data file is accessed by the Parametric Cost Estimating Model (PCEM) Computer Program in the generation of the facility types list and in breaking down the total estimated cost into the various categories of system cost. The name assigned to this data file, which is a simple text file, is “CostUniF.txt.”

```
Auditorium:             0.042:0.049:0.104:0.165:0.046:0.215:0.029:0.198:0.152:
College-Classroom:      0.025:0.023:0.148:0.070:0.025:0.210:0.016:0.299:0.158:
College-Dormitory:      0.025:0.015:0.175:0.129:0.018:0.252:0.023:0.207:0.115:
College-Laboratory:     0.081:0.039:0.068:0.076:0.043:0.223:0.000:0.358:0.097:
College-Student Union:  0.038:0.021:0.179:0.131:0.026:0.209:0.026:0.232:0.138:
Gymnasium:              0.050:0.045:0.182:0.110:0.039:0.215:0.000:0.239:0.105:
Hospital:               0.010:0.006:0.125:0.098:0.007:0.313:0.047:0.231:0.119:
Library:                0.037:0.022:0.189:0.166:0.025:0.167:0.032:0.250:0.112:
Office:                 0.026:0.019:0.133:0.127:0.020:0.249:0.039:0.239:0.148:
```

Appendix D PCEM Computer Program Codes

This Appendix contains the codes of the PCEM computer program.

DATA ENTRY

```
Private numCities As Integer
Private numBuildingTypes As Integer

Private privCityNames(0 To 100) As String
Private privCityFactors(0 To 100) As Double
Private privBuildingTypes(0 To 100) As String
Private privUniFactors(0 To 100) As Variant

Public Property Get CityNames() As Variant
  CityNames = privCityNames
End Property

Public Property Get CityFactors() As Variant
  CityFactors = privCityFactors
End Property

Public Property Get BuildingTypes() As Variant
  BuildingTypes = privBuildingTypes
End Property

Public Property Get UniFactors() As Variant
  UniFactors = privUniFactors
End Property

Private Sub cmbBuildingType_Change()
End Sub

Private Sub cmbProjectLocation_Change()
End Sub

Private Sub cmdExit_Click()
  End    'exit program
End Sub

Private Sub cmdProceed_Click()

  If txtInstitution = "" Or txtFacilityName = "" Or txtGrossSqFt = "" Or txtAssgSqFt = "" Or _
    txtNumFloor = "" Or txtEscFactor = "" Then
    MsgBox ("Please make sure to complete your input")
  Else
    If cmbProjectLocations.ListIndex = -1 Then
      frmReqCityInfo.Show
      If frmDataEntry.cmbBuildingTypes.ListIndex = -1 Then
        frmReqCityInfo.BuildingTypeInfoRequired = True
      End If
    ElseIf cmbBuildingTypes.ListIndex = -1 Then
      frmReqBuildingTypeInfo.Show
    Else
      frmModel.Show
```

```
      End If
      frmDataEntry.Hide
    End If
End Sub

Private Sub Form_Load()
    readCityFile
    readCostUnitFile
    initLists
End Sub

Private Sub initLists()
    For i = 0 To numCities - 1
      cmbProjectLocations.AddItem privCityNames(i)
    Next

    For i = 0 To numBuildingTypes - 1
      cmbBuildingTypes.AddItem privBuildingTypes(i)
    Next
End Sub

Private Function cityName(TextLine) As String
    colonPos = InStr(TextLine, ":")
    cityName = Left(TextLine, colonPos - 1)
End Function

Private Function cityFactor(TextLine) As Double
    colonPos = InStr(TextLine, ":")
    cityFactor = Right(TextLine, Len(TextLine) - colonPos)
End Function

Private Sub readCityFile()
  Dim TextLine
  Open "c:\PCEM\cityindex.txt" For Input As #1
  i = 0
  Do While Not EOF(1)   ' Loop until end of file.
    Line Input #1, TextLine   ' Read line into variable.
    privCityNames(i) = cityName(TextLine)
    privCityFactors(i) = cityFactor(TextLine)
    i = i + 1
  Loop
  numCities = i
  Close #1
End Sub

Private Function buildingType(TextLine) As String
    colonPos = InStr(TextLine, ":")
    buildingType = Left(TextLine, colonPos - 1)
End Function

Private Function UniFactor(TextLine) As Variant
```

```
  lastColonPos = InStrRev(TextLine, ":")
  lastSpacePos = InStrRev(TextLine, " ", lastColonPos)
  factorString = Mid(TextLine, lastSpacePos + 1, lastColonPos - lastSpacePos)
  i = 0
  startFactor = 1
  Dim factor(1000) As Double

  Do While startFactor < Len(factorString)
    currentColonPos = InStr(startFactor, factorString, ":")
    factor(i) = Mid(factorString, startFactor, currentColonPos - startFactor)
    startFactor = currentColonPos + 1
    i = i + 1
  Loop
  UniFactor = factor
End Function

Private Sub readCostUnitFile()
 Dim TextLine
 Open "c:\PCEM\CostUniF.txt" For Input As #1
 i = 0
 Do While Not EOF(1)
   Line Input #1, TextLine
   privBuildingTypes(i) = buildingType(TextLine)
   privUniFactors(i) = UniFactor(TextLine)
   i = i + 1
 Loop
 numBuildingTypes = i
 Close #1
End Sub

Private Sub Label1_Click(Index As Integer)

End Sub
```

REQUEST BUILDING TYPE

```
Private Sub lblCityName_Click(Index As Integer)

End Sub

Private Sub cmdProceed_Click()
  frmReqBuildingTypeInfo.Hide
  frmModel.Show
End Sub

Private Sub Form_Load()

End Sub
```

REQUEST CITY INFORMATION

```
Public BuildingTypeInfoRequired As Boolean

Private Sub Command1_Click()
End Sub

Private Sub cmdProceed_Click()
  frmReqCityInfo.Hide
  If BuildingTypeInfoRequired Then
    frmReqBuildingTypeInfo.Show
  Else
    frmModel.Show
  End If
End Sub

Private Sub Form_Load()

End Sub
```

MODEL

```
Private Sub cmdExit_Click()
  End
End Sub

Private Sub cmdGenReport_Click()

  Set fs = CreateObject("Scripting.FileSystemObject")
  Set reportFile = fs.CreateTextFile("c:\PCEM\EstimateReport.txt", True)
  reportFile.writeLine ("===== Input Information =====")
  reportFile.writeLine ("Institution:" + txtInstitution.Text)
  reportFile.writeLine ("Name of Facility Proposed:" + txtFacilityName.Text)
  reportFile.writeLine ("Type of Facility Proposed:" + txtFacilityType.Text)
  reportFile.writeLine ("Location of Project:" + txtProjectLocation.Text)
  reportFile.writeLine ("Gross Square Foot:" + txtGrossSqFt.Text)
  reportFile.writeLine ("Assignable Square Foot:" + txtAssgSqFt.Text)
  reportFile.writeLine ("Number of Floor Level:" + txtNumFloor.Text)
  reportFile.writeLine ("Escalation Factor:" + txtEscFactor.Text)
  reportFile.writeLine ("Location Factor:" + txtLocationFactor.Text)
  reportFile.writeLine ("===== Cost Estimates =====")
  reportFile.writeLine ("Estimated Dollar per GSF:" + txtEstDollarPerGSF.Text)
  reportFile.writeLine ("Estimated Dollar per GSM:" + txtEstDollarPerGSM.Text)
  reportFile.writeLine ("Estimated Total Building Cost:" + txtTotalBuildingCost.Text)
  reportFile.writeLine ("1.0 Foundations:" + txtUni(0).Text)
  reportFile.writeLine ("2.0 Substructure:" + txtUni(1).Text)
  reportFile.writeLine ("3.0 Superstructure:" + txtUni(2).Text)
  reportFile.writeLine ("4.0 Exterior Closure:" + txtUni(3).Text)
  reportFile.writeLine ("5.0 Roofing:" + txtUni(4).Text)
  reportFile.writeLine ("6.0 Interior Construction:" + txtUni(5).Text)
  reportFile.writeLine ("7.0 Conveying:" + txtUni(6).Text)
```

```
   reportFile.writeLine ("8.0 Mechanical:" + txtUni(7).Text)
   reportFile.writeLine ("9.0 Electrical:" + txtUni(8).Text)
   reportFile.Close
   response = MsgBox("Estimate report was generated and placed in C:\PCEM", vbOKOnly,
"PCEM")
End Sub

Private Sub Form_Load()

   ' names
   txtInstitution.Text = frmDataEntry.txtInstitution.Text
   txtFacilityName.Text = frmDataEntry.txtFacilityName.Text
   txtGrossSqFt.Text = frmDataEntry.txtGrossSqFt.Text
   txtAssgSqFt.Text = frmDataEntry.txtAssgSqFt.Text
   txtNumFloor.Text = frmDataEntry.txtNumFloor.Text
   txtEscFactor.Text = frmDataEntry.txtEscFactor.Text

   If frmDataEntry.cmbProjectLocations.ListIndex <> -1 Then
     txtProjectLocation.Text = frmDataEntry.cmbProjectLocations.Text
     cityNumber = frmDataEntry.cmbProjectLocations.ListIndex
     txtLocationFactor.Text = frmDataEntry.CityFactors(cityNumber)
   Else
     txtProjectLocation.Text = frmReqCityInfo.txtCityName
     txtLocationFactor.Text = frmReqCityInfo.txtCityFactor
   End If


   ' Calculate estimates
   txtEstDollarPerGSF.Text = Format((202.245 + (15.74 *
Log(CDbl(frmDataEntry.txtNumFloor))) - _
                              (126.196 * CDbl(frmDataEntry.txtAssgSqFt) / _
                              CDbl(frmDataEntry.txtGrossSqFt))) * _
                             CDbl(txtLocationFactor.Text) * _
                             CDbl(txtEscFactor.Text), "###,##0.00")

   txtEstDollarPerGSM.Text = Format(CDbl(txtEstDollarPerGSF.Text) / 0.093, "###,##0.00")
   txtTotalBuildingCost.Text = Format(CDbl(txtEstDollarPerGSF.Text) *
CDbl(txtGrossSqFt.Text), _
                             "###,###,##0")
   ' 4 is the number of significant digits
   txtTotalBuildingCost.Text = Format(SigniRound(CDbl(txtTotalBuildingCost.Text), 4),
"###,###,##0")



   ' Calculate uniformat costs
   UniFactors = frmDataEntry.UniFactors
   buildingTypeNumber = frmDataEntry.cmbBuildingTypes.ListIndex
   If buildingTypeNumber <> -1 Then
     txtFacilityType.Text = frmDataEntry.cmbBuildingTypes.Text
     For i = 0 To 8
```

```
        txtUni(i).Text = Format(CDbl(txtTotalBuildingCost.Text) * _
                         CDbl(frmDataEntry.UniFactors(buildingTypeNumber)(i)),
"###,###,###,##0")
        txtUni(i).Text = Format(SigniRound(CDbl(txtUni(i).Text), 4), "###,###,###,##0")
      Next
    Else
      txtFacilityType.Text = frmReqBuildingTypeInfo.txtFacilityType.Text
      For i = 0 To 8
        txtUni(i).Text = Format(CDbl(txtTotalBuildingCost.Text) * _
                         CDbl(frmReqBuildingTypeInfo.txtUni(i).Text), "###,###,###,##0")
      Next
    End If
End Sub


' Round the given number to the specified significant number of digits
Function SigniRound(num As Double, signi_digits As Integer) As Double
    Dim divisor As Long
    Dim result As Double
    divisor = 1
    result = 2
    Do While result > 1
      divisor = divisor * 10
      result = num / divisor
      Debug.Print "result = " + CStr(result)
    Loop
    divisor = divisor / 10
    Debug.Print "divisor = " + CStr(divisor)
    num = num / divisor
    num = Round(num, signi_digits - 1)
    Debug.Print "rounded divided num = " + CStr(num)
    SigniRound = num * divisor
End Function
```

Bibliography

Al Khalil, M., Assaf, S., Rahman, W. A., and Asfoor M., A Conceptual Cost Estimating Model for Water Reservoirs, Cost Engineering, Vol. 41, No. 5, May 1999, p. 38-43.

Al Tabtabai, H., Alex, A. P., and Tantash, M., Preliminary Cost Estimation of Highway Construction Using Neural Networks, Cost Engineering, Vol. 41, No. 3, March 1999, p. 19-24.

AACE Recommended Practice No. 17R-97: Cost Estimate Classification System, AACE, Inc., 1997.

Babbie, Earl R., The Practice Of Social Research, 6th Edition, Wadsworth Pub. Co., CA, 1992.

Barrie, D. S., and Paulson, B. C., Professional Construction Management, McGraw-Hill, NY, 1992.

Black, James H., Application of Parametric Estimating to Cost Engineering, 1984 AACE Transactions, Paper B.10, 1984.

Bley, A. F. S., Improved Conceptual Estimating Performance Using A Knowledge-Based Approach, Dissertation presented to the University of Texas at Austin in partial fulfillment of the requirements for the degree of Doctor of Philosophy, May, 1990.

Bode, Jürgen, Neural Networks for Cost Estimation, Cost Engineering, Vol. 40, No. 1, January 1998, p. 25-30.

Bowlby, Roger L., and Schriver, William R., Observations on Productivity and Composition of Building Construction Output in the United States, 1972-82, Construction Management and Economics, Vol. 4, No. 1, p. 1-18, 1986.

Carr, Robert I., Critique - Cost Estimating and Project Management, Proceedings of the Conference on Current Practice in Cost Estimating and Cost Control, ASCE, Austin, Texas, 1983, p. 61-67.

Carr, Robert I., Cost-Estimating Principles, Journal of Construction Engineering and Management, Vol. 115, No. 4, December 1989, p. 545-551.

Cho, Chung-Suk, Development of The Project Definition Rating Index (PDRI) For Building Projects, Dissertation presented to the University of Texas at Austin in partial fulfillment of the requirements for the degree of Doctor of Philosophy, May, 2000.

The Construction Industry Institute, Pre-Project Planning: Beginning a Project the Right Way - CII Publication 39-1, Austin, Texas, Construction Industry Institute, The University of Texas at Austin, 1994.

The Construction Industry Institute, Pre-Project Planning Handbook – Special Publication 39-2, Austin, Texas, Construction Industry Institute, The University of Texas at Austin, 1995.

Collier, Keith, Estimating Construction Costs: A Conceptual Approach, Reston, Reston, VA, 1984.

Creese, R., and Li, L., Cost Estimation of Timber Bridges Using Neural Networks, Cost Engineering, Vol. 37, No. 5, 1995, p. 17-22.

Daschbach, J. M. and Apgar, H., Design Analysis Through Techniques of Parametric Cost Estimation, Engineering Costs and Production Economics, 14, p. 87-93, 1988.

Dell'Isola, Alphonse J. and Kirk, Stephen J., Life Cycle Costing for Design Professionals, McGraw-Hill, NY, 1981.

Di Natale, J., Conceptual Estimating Requirements and Conditions, Transactions of the American Association of Cost Engineers, Paper B-5, 1980.

Dysert, Larry R., Developing a Parametric Model for Estimating Process Control Costs, Cost Engineering, Vol. 43, No. 2, February 2001, p. 31-34.

Ferry, Douglas J. and Brandon, Peter S., Cost Planning of Buildings, 6th Edition, Billing & Sons, Worcester, Great Britain, 1991.

Garza, Jesus M. and Rouhana, Khalil G., Neural Networks Versus Parameter-Based Applications in Cost Estimating, Cost Engineering, Vol. 37, No. 2, February 1995, p. 14-18.

Gibson, G. E., Kaczmarowski, J. H., and Lore, H. E., Modeling Pre-Project Planning for the Construction of Capital Facilities - Source Document 94, Austin, Texas, Construction Industry Institute, The University of Texas at Austin, 1993.

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

Hamilton, Lawrence C., Regression with Graphics: A Second Course in Applied Statistics, Duxbury Press, CA, 1992

Hegazy, Tarek, and Ayed, Amr, Neural Network Model for Parametric Cost Estimation of Highway Projects, Journal of Construction Engineering and Management, ASCE, Vol. 124, No. 3, May/June 1998, p. 210-218.

Hegazy, T., Moselhi, O., and Fazio, P., Developing Practical Neural Network Application Using Backpropagation, Journal of Computing in Civil Engineering, Vol. 9, No. 2, p. 145-159, 1994.

Karshenas, Saeed, Predesign Cost-Estimating Method for Multistory Buildings, Journal of Construction Engineering and Management, ASCE, Vol. 110, No. 1, March 1984, p. 79-86.

Knoke, David, and Bohrnstedt, George, Statistics for Social Data Analysis, 3rd Edition, F.E. Peacock Publishers, Inc., IL, 1994.

Kouskoulas, Vasily, and Koehn, Edward, Predesign Cost-Estimation Function for Buildings, Journal of the Construction Division, ASCE, Vol. 100, No. CO4, December 1974, p. 589-604.

Kreps, Rollin A., and Slomba, John W., Conceptual Estimating Systems and Their Benefits, 1990 AACE Transactions, Paper K.2, 1990.

Krieg, Tom E., An Introduction to Parametric Estimating, 1979 Transactions of the American Association of Cost Engineers, Paper F-3, 1979.

Logcher, R. D., Techniques for Forecasting Costs, ASCE Specialty Conference on Building Economics, New Hampshire, 1980.

Lorance, Randal B., and Wendling, Robert V., Basic Techniques for Analyzing and Presenting Cost Risk Analysis, 1999 AACE International Transactions, Paper RISK.01, 1999.

Ludwig, Ernest E., Applied Project Engineering And Management, 2nd Edition, Gulf Publishing Co., 1988.

Means Company, Inc., Building Construction Cost Data 2002, 60th Edition, Waier, Philip, Editor, R.S. Means Co., MA, 2001.

Means Company, Inc., Means Square Foot Cost 2002, 23rd Edition, Waier, Philip, Editor, R.S. Means Co., MA, 2001.

Melin, John B. Jr., Parametric Estimation, Cost Engineering, Vol. 36, No. 1, January 1994, p. 19-24.

Meyer Edward R., and Burns, Thomas J., Facility Parametric Cost Estimating, 1999 AACE International Transactions, Paper EST.02, 1999.

Mitchell, Frank B. Jr., Developing a Project Scope/Conceptual Estimating Questionnaire, AACE Transactions, 1995.

Moselhi, O., Hegazy, T., and Fazio, P., Potential Applications of Neural Networks in Construction, Canadian Journal of Civil Engineering, Vol. 19, p. 521-529, 1992.

Noble, William, Conceptual Estimation and Budget Control, 1987 AACE Transactions, Paper C-11, 1987.

Norušis, Marija J., SPSS® 6.1 Guide to Data Analysis, Prentice-Hall, NJ, 1995.

O'Connor, Patrick D., Practical Reliability Engineering, 2nd Edition, John Wiley & Sons, Inc., 1985.

Orczyk, Joseph, and Chang, Luh-Maan, Parametric Project Scheduling of Milestones, AACE Transactions, 1990.

Ostwald, Phillip F., Construction Cost Analysis and Estimating, Prentice-Hall, Inc., 2001.

Park, Chansik, Choi, Seok-In, and Kim, Ki-Hong, An Efficient Quantity Takeoff Method for the Preliminary Design Phase, 1999 AACE International Transactions, Paper EST.09, 1999.

Parker, Donald E., Budgeting By Criteria Not Cost Per Square Foot, 1984 AACE Transactions, Paper A-3, 1984.

Parker, Donald E., Project Budgeting for Buildings, 1993 AACE Transactions, Paper C.14, 1993.

Popham, Kymberli, Cost Estimating Using Historical Costs, 1996 AACE Transactions, Paper EST.14, 1996.

PMI Standards Committee, A Guide To The Project Management Body Of Knowledge, Project Management Institute, Newtown Square, PA, 1996.

Rose, Aaron, An Organized Approach to Parametric Estimating, Transactions of the Seventh International Cost Engineering Congress, London, England, October, 1982.

Skitmore, Martin, Factors Affecting Accuracy of Engineers' Estimates, 1988 AACE Transactions, Paper B-3, 1988.

Smith, M.A., and Tucker, R.L., An Assessment of the Potential Problems Occurring in the Engineering Phase of an Industrial Project. A Report to Texaco, Inc., The University of Texas at Austin, 1983.

Stevens, James, Applied Multivariate Statistics for the Social Sciences, 3rd Edition, Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

Stewart, R. D., "Fundamentals of Cost Estimating", in Cost Estimator's Reference Manual, edited by R. Stewart and R. Wyskida, John Wiley & Sons, Inc., NY, 1987.

Tabachnick, Barbara G. and Fidell, Linda S., Using Multivariate Statistics, 4th Edition, Allyn and Bacon, Boston, MA, 2001.

Uppal, K. B., Process Cost Engineering in the 21st Century, 2000 AACE International Transactions, Paper CSC.02, 2000.

Vita

Kan Phaobunjong was born in Chachoengsao, Thailand on November 22, 1971, the son of Thiam and Raynu Phaobunjong. At the age of ten, he left Thailand to pursue his study in Singapore. After completing his study at Anglo-Chinese Junior College, Singapore, in 1991, he returned to Thailand and worked as a supervisor of a chicken farm. In June 1992, he began his university education at Lamar University, in Beaumont, Texas. In January 1994, he transferred to The University of Texas at Austin. He received the degree of Bachelor of Science in Civil Engineering from The University of Texas at Austin in December 1995. From February to July 1996, he worked as a structural engineer and assistant to the project coordinator for Line Consultants Co., Ltd., in Bangkok, Thailand. In August 1996, he entered The Graduate School at The University of Texas at Austin and was awarded the degree of Master of Science in Civil Engineering in December 1997. He subsequently entered into the Ph.D. program in civil engineering at The University of Texas at Austin in 1998. During that time, he was a teaching assistant for the construction cost estimating class and construction surveying class. He also co-authored a textbook, *Estimating Building Cost*, with Dr. Calin M. Popescu and Nuntapong Ovararin. After four and a half years of study and research, he will graduate in 2002 with the degree of Doctor of Philosophy with a major in civil engineering.

Permanent address: 31/3 Moo 6, Tarart Klongsuan,
Tumbon Teaparart, Aumpur Banpoe,
Chachoengsao 24140, Thailand.

This dissertation was typed by the author.