1988

Parametric estimating: an equation for estimating buildings.

McGarrity, Robert J.

PARAMETRIC ESTIMATING: AN EQUATION FOR ESTIMATING BUILDINGS

A Special Research Problem

Presented To

The Faculty of The School of Civil Engineering

by

Robert J. McGarrity

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Civil Engineering

Georgia Institute of Technology

September, 1988

Approved:

____

Faculty Advisor/Date

____

Reader/Date

____

Reader/Date

ABSTRACT

Historically, the need for accurate and reliable cost estimates prior to the actual design has proven to be invaluable. One technique being utilized by the construction industry to fulfill this need is parametric estimating. The objective of this paper is to develop a parametric estimating model. In order to achieve this goal the concepts and theory behind parametric estimating are first explained and then demonstrated by the presentation of two previously published parametric models. Lastly, a parametric model developed to provide predesign estimates for buildings is explained and tested.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 INTRODUCTION

In today's world of shrinking budgets and complicated financing schemes the first and most reoccurring question asked by owners to their design personnel is **HOW MUCH WILL IT COST?** In an attempt to accurately answer this question many techniques have been developed and used to predict the cost of a project prior to the actual design. One technique being utilized by the construction industry is parametric estimating. In this paper the concepts and methodology behind parametric estimating will be described. Additionally, with the intent of giving the reader a better understanding for processes involved in model development, two separate parametric models that have previously been published in the ASCE'S Journal of the Construction Division will be presented and critique. Following this, the remainder of the paper will be dedicated to the discussion of a parametric cost-estimating model that was developed from data collected from actual construction projects. This model and the procedures used in it's formulation will be discussed in detail. Lastly the accuracy of the developed model will be tested by its ability to predict the cost of two buildings whose actual cost figures are known.

1.2 PROBLEM STATEMENT

The need for an accurate and reliable cost estimate prior to the actual project design has historically been essential to the success of all construction projects. A look at the formative stages in the building process, Figure 1.1, reveals that a project is proposed for construction in an effort to fulfill an identified need. This recognition of a need is the first step in the building process [Halpin-Woodhead 80]. From this need a project is conceptualized. At this stage in the process a decision as to whether or not it is feasible to proceed along down the building process line must be made. This decision process is commonly referred to as a feasibility analysis. Although any sound feasibility analysis considers all pertinent factors relevant to the project, the initial estimate as to the total project cost is normally the most weighted factor used in making the decision. As such, the initial estimate is used to screen and eliminate unsound proposals and decide whether money should be invested so that the project may proceed to the next step in the process, design.

Although the value of an accurate predesign estimate is enormous, it is usually performed without the benefit of: detailed drawings and specifications, knowledge of what construction methods are to be employed, time, and money. Therefore, it is essential that fast, inexpensive, and reasonably accurate methods to estimate a project, before the detailed plans

and specifications are prepared, be explored and developed [Karshenas 84].

Consequently, the primary objective of this research is to investigate the practicality and usefulness of developing a predesign parametric estimating model based on historical cost data.

1.3 PROCEDURE

In an effort to efficiently accomplish the above objective, the concepts of parametric estimating will first be discussed and explained. This introduction to parametric techniques will be followed by a presentation and analysis of two previously published parametric models. Lastly, a parametric model developed as the culmination of this research will be described and tested.

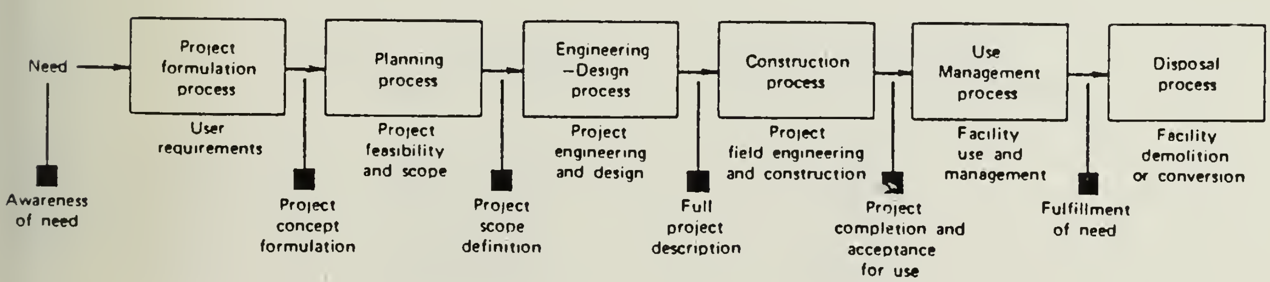

FIGURE 1.1 The Building Process [Halpin-Woodhead 80]

CHAPTER 2 PARAMETRIC ESTIMATING

2.1 INTRODUCTION

In this part of the text the concept of parametric estimating will be introduced, defined and illustrated. Additionally, the steps involved in the successful development of a parametric cost-estimating model will be identified and discussed individually.

2.2 PARAMETRIC ESTIMATING DEFINED

Parametric estimating is the process of estimating cost by using mathematical equations that relate cost to one or more physical or performance variables associated with the item being estimated [Wyskida-Steward 87]. Used in its most simplest form, a unit estimate that predicts the cost of a building based on its square footage is a parametric estimate as it relates the cost of the building to one physical variable - the square footage. As an example of a unit cost estimate consider the following fictitious cost data:

	TOTAL BUILDING COST ($)	TOTAL BUILDING SQUARE FOOTAGE (sf)
PROJECT 1	100,000	2,000
PROJECT 2	145,000	3,000
PROJECT 3	190,000	4,000
PROJECT 4	225,000	5,000
Totals	660,000	14,000

Table 2.1 Fictional Building Costs

From Table 2.1, the unit cost of a typical square foot for a building can easily be calculated by:

UNIT COST = $660,000/14,000 sf = $47.14 per sf

Most often this is the way that unit prices are derived. Of course the sample size used by published estimating manuals is far greater than four, but in principal the procedure is the same.

The obtained value of $47.14/sf can now be multiplied by the total area of any proposed building to obtain an estimate of the building's cost. This simple technique of multiplying the square footage of a building by a unit cost is the most popular of all preliminary estimating techniques [Ostwald 84].

A unit cost estimate is easily converted to a parametric one by simply expressing the cost in the form of an equation in which cost is related and dependent on one or more physical or performance variables. In the case of the above example it can easily be fitted to an equation of the form:

$$C = 47.14 * A \qquad (2.1)$$

where, C = Cost of the proposed building and is termed a dependent variable since its value is dependent on that of A.

A = The number of square feet in the proposed building and is termed an independent variable since its value does not dependent on another variable.

47.15 is a parametric value based on the historical data from Table 2.1 and is used to relate the dependent variable (cost) to the independent variable (square footage).

2.3 ORIGINS OF PARAMETRIC ESTIMATING TECHNIQUES

The first documented uses of the application of statistical techniques to modeling occurred in the late 1950's and was initiated and pursued by the Rand Corporation in an attempt to predict military hardware cost at very early phases of design. Its use by Rand to obtain credible cost estimates, while projects were still in the conceptual design phase, drew much attention from both Government and the private sector. Both communities were quick to recognize the derived benefits of having early estimates that were not time and labor intensive like previous detailed techniques.

Through the years, the fields of business, macroeconomics and social science have used parametric estimating as a means of

correlating observations of past events and occurrences to predict future happenings. Recently, the proliferation of computers and software has simplified the chore of maintaining a data base and performing complicated statistical calculations and analysis. As a result, today parametric estimating is being used to some degree in all fields where cost estimating and forecasting take place. Professional organizations whose members are involved in parametric estimating include the American Association of Cost Engineers, the International Society of Parametric Analysts, the National Estimating Society, and the Institute of Cost Analysis.

2.4 DEVELOPMENT OF A PARAMETRIC COST ESTIMATING MODEL

Although there appears to be no set algorithm for the development of an estimating model of this kind, review of past work in this area as well as general readings on the subject have revealed four reoccurring steps that appear to be essential to the successful development of a parametric model. The four steps in the order in which they should be performed are:

1. Parameter Selection
2. Data Collection and Normalization
3. CER Form Selection and Derivation
4. Measuring the Goodness of Fit/Model Testing

The remainder of this chapter will be dedicated to the discussion of the above four steps.

2.4.1 PARAMETER SELECTION

As stated in section 2.2, a parameter is a physical or functional characteristic upon which the total cost of the project is largely dependent, (for the purposes of this paper the project will be the construction of a building). Sometimes these characteristics or parameters are called "cost drivers" as they should be highly correlated with cost [Ostwald 84]. From historical data, empirical coefficients are determined and fitted in a cost equation. These cost equations that are used to model the cost function are known as cost-estimating relationships (CERs). After the development of a CER the actual value of the physical characteristics of a proposed project are obtained from the design, substituted into the CER and the estimated cost calculated. Some examples of parameters applicable to building construction are:

Gross Floor Area

Roof Area

Length of Building Perimeter

Height of Building.

Number of Floors

The success of any parametric model is dependent upon what parameters are utilized. Some factors that should be considered prior to selecting "cost drivers" are:

1. As stated, the characteristic should be highly correlated to the building cost.

2. The developer must be assured that data concerning the parameter is available from past projects. Additionally, if the derived model is to be used as a predesign estimating tool, the actual values for the parameter or at least a rough estimate must be obtainable prior to actual design.

3. If the goal of the model is to develop a fast efficient way to estimate the cost of a building, then the number of parameters used should be kept at a minimum, with only those characteristics necessary to define the essential cost components of the building being used.

2.4.2 DATA COLLECTION AND NORMALIZATION

Once the parameters have been decided upon, the next step in the model development process is data collection. This step, although at times very tedious, is essential as a model's ability to predict future costs based upon historical data is totally dependent on the data base from which it was derived. As a

consequence, standard ground rules should be developed so that all data is collected in the same manner. In the case of developing a model to estimate building costs, plans and specifications will have to be reviewed and the necessary values of the physical characteristics chosen as parameters obtained. After the values of the parameters are obtained cost data for the project must be collected. This data must be similar for all projects. That is to say, that the cost for certain non-common items must be excluded from the total cost. For example, if all buildings do not have shallow foundations (but the majority do), the cost differences between the installed deep foundations and what a shallow foundation would have cost should be subtracted from the total cost of that particular building. Similarly, rules need to be established for overhead, profit, and all other similar items. In short, the key to good data collection is consistency. The cost of certain items must be added or deleted to a project to make it identical to the rest. This process is called normalization and is essential to the model building process. In addition to the above mentioned items, data must be normalized for location, year built, quality and any other factors that might differentiate a project from the norm. At the completion of the normalization process one is left with a data base containing buildings whose cost can accurately be compared to one another.

2.4.3 SELECTION AND DERIVATION OF THE PROPER CER

2.4.3.1 COST-ESTIMATING RELATIONSHIPS

The end product of any parametric model is the cost-estimating relationship. Although numerous possible mathematical equation forms can be used for a cost-estimating relationship, most cost data can be fit empirically using one of the forms shown in Figure 2.1.

As discussed previously, the simplest CERs are no more complicated than the unit cost example from section 2.2. Linear relationships similar to equation 2.1 are of the form Y = AX. Note that use of a form like equation 2.1 represents the situation where no fixed costs are present (e.g., Land Purchase, Mobilization, etc.). That is to say that when no square footage of a building is built the cost is 0 dollars. Another limitation of the use of a linear equation of this form is that it fails to account for the economies of scale inherent in the construction industry. In short the principal of economies of scale as it pertains to construction says that, in general, a large building should cost less per square foot than a small one [Wyskida-Steward 87].

One improvement to the basic CER expressed by equation 2.1 is the use of another linear equation form, namely: Y = A + BX. Use of this form indicates the presence of both fixed and variable costs. The fixed costs component is represented by the

A term while the variable cost component is the BX term. An equation of this form may be obtained from the data in Table 2.1 by performing a simple linear regression utilizing the method of least squares. The results of this regression analysis yield the following equation:

$$C = 42A + 18,000 \qquad (2.2)$$

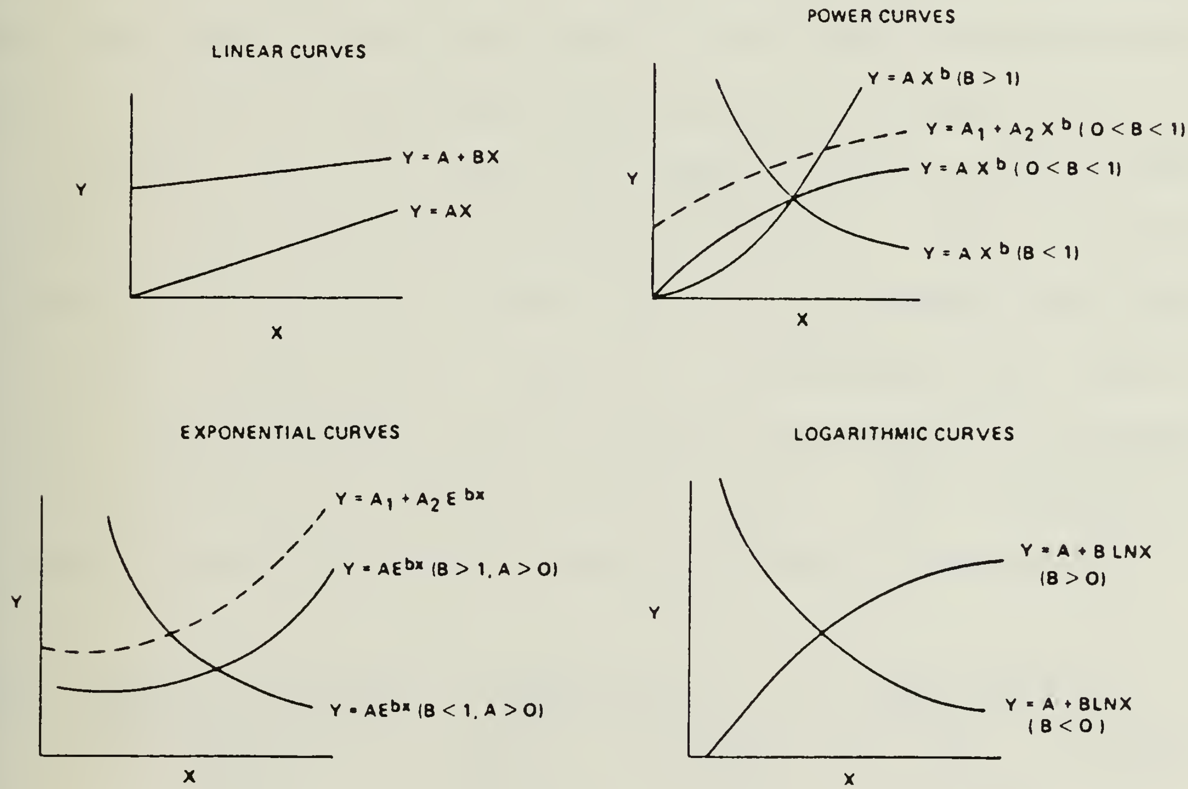

Figure 2.1 CER Function Shapes on Linear Coordinates

[Wyskida-Steward 87]

Note that the regression equation has a positive y-intercept, which from a practical standpoint makes sense since it represents positive fixed costs. On the other hand, fixed cost values less than zero are an unlikely situation and are usually indicative of faulty data or that the equation model is suspect.

The use of linear CERs in the form of equations 2.1 and 2.2 guarantees that a change of one unit in the independent variable ,A, will be accompanied by a constant change in the dependent variable ,C, as determined by the coefficients of the particular equation. As previously discussed, equation 2.1 fails to take advantage of economies of scale and thus estimates the same dollar per square foot regardless of the size of the building. Equation 2.2 on the other hand, does improve on equation 2.1 in this respect, as its cost per square foot of building does decrease as the size of the building increases. This unit cost reduction is the result of the fixed cost additive term being spread over a larger area. Table 2.2 illustrates this fact by showing the calculated total and unit costs for four proposed buildings using both equation 2.1 and 2.2. The principle of economies of scale will be discussed more fully in Chapter 3.

BLDG. #	PROPOSED BUILDING SQUARE FOOTAGE (SF)	EQUATION 2.1 ESTIMATE ($)	EQUATION 2.1 UNIT COST ($/SF)	EQUATION 2.2 ESTIMATE ($)	EQUATION 2.2 UNIT COST ($/SF)
1	2000	94,280	47.14	102,000	51
2	3000	141,420	47.14	144,000	48
3	4000	188,560	47.14	186,000	46.5
4	5000	235,700	47.14	228,000	45.6

TABLE 2.2 Demonstration of Equations 2.1 and 2.2

Unfortunately, the variable costs associated with most construction projects usually do not behave in a linear manner. In particular, it is common for the economies of scale in the construction industry to be such that not only does the cost per square foot of a building decrease with size, but also so does the unit increment of variable cost. Models exhibiting these characteristics are of course non-linear and normally take the form of power curves, exponential curves, or logarithmic curves as shown in Figure 2.1.

The use of a power curve assumes a relationship between the independent variables and cost, such that a percentage change in the independent variables causes a relatively constant percentage change in cost. The inset in Figure 2.2 demonstrates this by showing that for these particular power curve coefficients, successive 50% changes in the independent variable cause successive 25% changes in cost. For a pure power curve in the form $Y = AX^b$, the percent change in the dependent variable is a constant percentage, whereas for a power curve of the form $Y = A_1 + A_2{}^2X^b$, the change in the dependent variable will depart from a constant percentage depending on the relative magnitude of the A_1 term [Wyskida-Steward 87].

As with the power curve the exponential CER may or may not have an additive term. Use of this form however, assures a relationship between the independent variable and cost such that a unit change in the independent variable causes a relatively constant percentage change in cost. This is shown in Figure 2.3

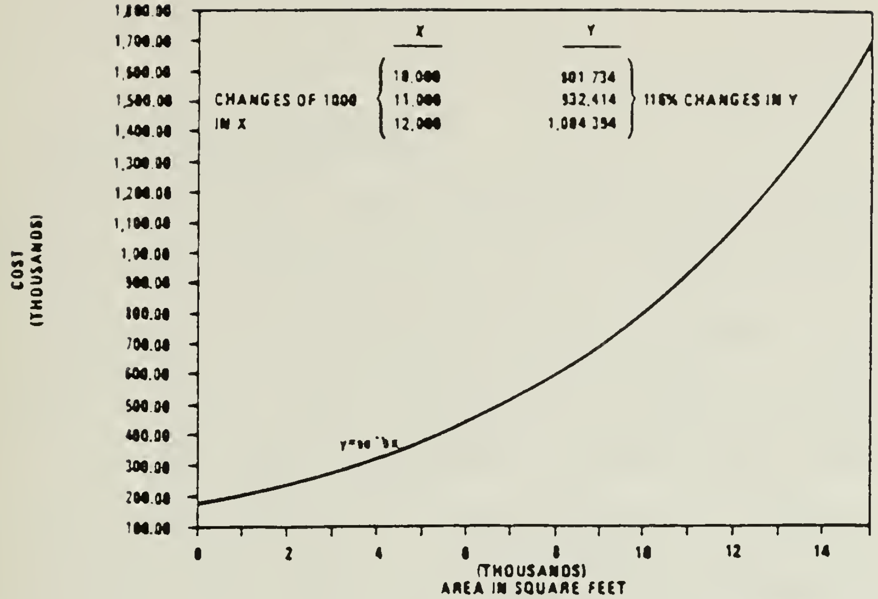

FIGURE 2.2 Power Function CER [Wyskida-Steward 87]

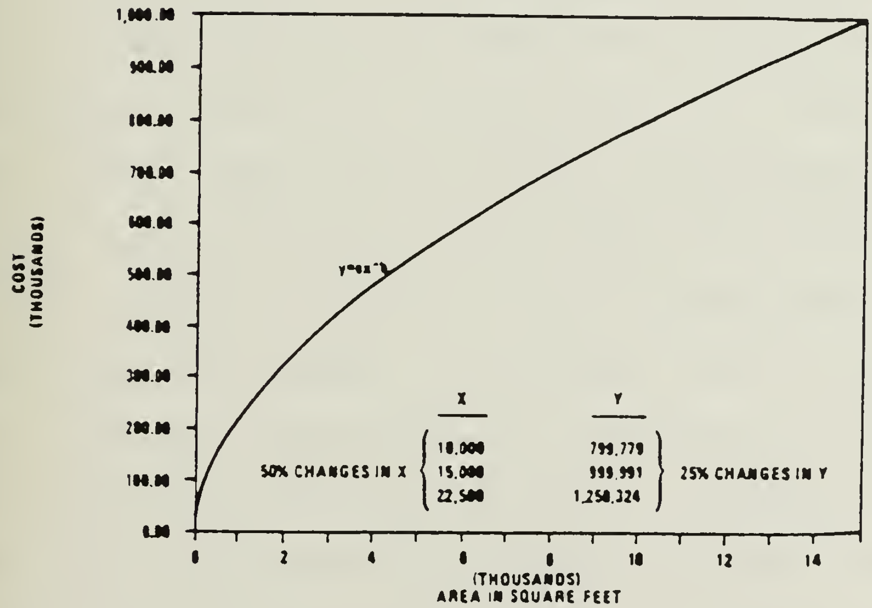

FIGURE 2.3 Exponential Function CER [Wyskida-Steward 87]

where, for $Y = Ae^x$ successive 1000 unit changes in x cause successive 116% changes in cost.

2.4.3.2 CER SELECTION

The choice of which form from Figure 2.1 to use for a given set of data can be accomplished by one of three possible methods. Method 1 which is non-mathematical is based on the users understanding of the cost-estimating relationships previously described. A thorough knowledge of what each form physically represents as well as the limitations of each, allows the experienced model developer to chose the CER form to match his data and the particular circumstance being modeled.

Method 2, on the other hand, is a graphical technique in which the best equation form can be determined by discovering what kind of graph paper, (linear-linear, linear-logarithmic, or logarithmic-logarithmic), that best permits a straight line to be drawn through a scatter plot of the data. If plotting the data on linear-scaled graph paper produces a data pattern that can be fitted well with a straight line the best fit CER will be of a linear form either $Y = AX$ or $Y = A + BX$ depending on whether or not fixed costs are present. If the best fit on linear paper is a curve, then the data should be replotted on semi-logarithmic paper and the best fit line redrawn. A straight line here indicates that the best CER form will be either a exponential or logarithmic equation. If the best fit on semilog paper is a curve, replot the data on full log paper. A straight line here

indicates that a power curve is the most appropriate cost-estimating form. In the last two instances, a slight curve on either semilog paper or log-log paper may be correctable by the addition of a constant to the CER equation [Wyskida-Steward 87].

The third method for determining which equation to use for the form of the CER is purely mathematical and involves using results obtained from multiple regression analysis. In particular, the coefficient of multiple determination defined as:

$$\text{R-squared} = R^2 = \frac{\sum_{i=1}^{n} (C_i - C_{average})^2}{\sum_{all} (C_{actual} - C_{average})^2} \quad .....(2.3)$$

where, C_i represents the predicted cost of $project_i$ using the derived CER, C_{actual} is the actual cost of building i, $C_{average}$ is the average cost of all the buildings used in the sample and n is the total number of projects in the sample data.

The value of R-squared is a measure of the closeness of fit of the regression equation to the observed points. An R-squared value of 1 would indicate that the selected form and the derived equation for the CER perfectly predicts the building cost. Therefore, to use the R-squared value as a criterion for CER selection multiple regression analysis must first be performed to fit the normalized data to each of the possible CER form candidates . After this is done and the R-squared values for each CER are calculated, and CER selection can be accomplished by simply choosing the form with the R-squared value closest to 1. The use of the R-squared value in model development will be discussed more fully in Chapter 3.

2.4.3.3 CER DERIVATION

After determining the equation form that is best suited for the data, the next step is to derive the mathematical equation for the CER (of course if method 3 from Section 2.4.3.2 is employed as a means of selecting the CER form, then a mathematical equation has already been derived). To accomplish this task, statistical methods of multiple regression are employed. Although graphical and hand algebra techniques do exist to perform multiple regressions in which several independent variables are related to an dependent variable, they have largely been replaced by computer programs that use various statistical methods to quickly and efficiently derive CER equations. The most common of these methods used is called the "method of least squares" and the reader is referred to any college statistics text book for its derivation and use.

2.4.4 MEASURE OF GOODNESS OF FIT/MODEL TESTING

Several statistical criteria and variance analysis techniques are used to measure the goodness of fit for any regression analysis. Two of the more common techniques are the R^2 value previously discussed and the standard error (S.E.).

The standard error measures the average amount by which the actual costs differs from the calculated costs by:

$$S.E. = \frac{\sum_{i=1}^{n} (C_{actual} - C_i)^2}{n - 1} \quad \ldots\ldots\ldots\ldots\ldots\ldots(2.4)$$

where the variables are defined as in equation 2.3. Because it is desirable to have a cost-estimating relationship that produces calculated values that are very close to the actual costs, the smaller the standard error the better. Notice that the units of standard error are the same as C_i and $C_{calculated}$ and for the purposes of this paper, as we are dealing with costs, it will always be dollars ($).

In addition to the two above statistical techniques for measuring the goodness of fit, the calculated residual values, $(C_{actual}-C_{predicted})$, or the percent residuals, $(C_{actual}-C_{predicted}/C_{actual})$, may be used as a feel for the accuracy of any regression model. More practically, another way to measure the appropriateness of any cost-estimating model, is to put it to use and test the validity of its results. One way to do this is to simply use the derived model to estimate a sample of actual projects whose cost data are known but were not used in the actual development of the CER. A comparison of these project's actual cost with that of their estimated values may be used as an indicator of the accuracy that can be expected from the derived model.

CHAPTER 3 BACKGROUND WORK

3.1 INTRODUCTION

Although the practice of parameter estimating has been in existence for over three decades, the use and acceptance of it by the construction industry as a valid estimating technique is a relatively recent event. As a result, much has been written on the topic of parametric estimating in general, however, little has been published having to do with the methodology involved in the actual development of a parametric estimating model that pertains to construction. Two such articles that do exist were published in the American Society of Civil Engineers Journal of the Construction Division. The first of the articles appeared in December of 1974 and was written by V. Kouskoulas and E. Koehn and was entitled "Predesign Cost Estimating Function for Building". The second article on model development was published in March of 1984 by Saeed Karshenas and was titled "Predesign Cost Estimating Method for Multistory Buildings". Collectively, the principles and procedures established by these two papers form the foundation upon which the model presented in Chapter 4 of this paper was built.

In this chapter, each of the two authors' papers will be summarized and presented. In the section immediately following the explanation of both papers, each will be critique in an effort to point out obvious weaknesses.

3.2 SUMMARY OF V. KOUSKOULAS AND E. KOEHN'S MODEL

In their article the authors use a three phase approach to the problem of model development [Kouskoulas-Koehn 75]. The first of the phases deals with the selection of the independent variables upon which the cost of a building depends. After the variables have been established, the second phase of the model development is to chose the appropriate form of the cost-estimating relationship that will properly relate the selected independent variables to that of the dependent cost variable. After the form is selected the mathematical relationship must then of course be derived. Lastly, after the cost function has been derived it must be tested as to its reliability and acceptability for use. Taken collectively, the authors' discussion of their rational and reasoning in each of these phases make up their paper. As such, each of these areas will be discussed in the following pages.

3.2.1 THE SELECTION OF THE INDEPENDENT VARIABLES

The authors state three basic criteria that they used to select their independent variables. These criteria are listed below:

1. The variable must physically describe the project in some way, while also being a major contributor to the total cost.

2. The data for each variable must be available and retrievable from both completed projects and for future proposed projects.

3. The variables must be varied and general enough so that any function derived with their use would be applicable to a wide class of building projects at any moment in time and for any place.

Guided by these three criteria the authors selected and defined six independent variables. In their opinion, these six variables are specific enough to adequately describe the building while at the same time general enough to define a "global predesign cost estimating function". The particular variables describe the building by its location, time of realization, function or type, height, quality, and technology. A description of the variables, with the reasoning behind each, is given below:

1. C was selected to be the dependent variable representing the cost. In this article however the C does not represent a total cost but rather a unit cost for the building. That is to say, that the value derived for C will have a dollars per square foot term as its units.

2. The locality variable, L, identifies the differences in construction costs as a consequence of differences in the style and cost of living between different cities as well as wage differentials resulting from differences in labor structures. The accompanying Table 3.1 was provided as one possible source for a locality index.

City (1)	Index, *L* (2)
Boston, Mass.	1.03
Buffalo, N.Y.	1.10
Dallas, Tex.	0.87
Dayton, Ohio	1.05
Detroit, Mich.	1.13
Erie, Pa.	1.02
Houston, Tex.	0.90
Louisville, Ky.	0.95
New York, N.Y.	1.16
Omaha, Neb.	0.92

TABLE 3.1 Locality Index [Kouskoulas and Koehn 74]

3. The price index variable, P, is time dependent and used by the authors to predict the future price indexes from historical data. From the data provided in Table 3.2, the following expression was developed to define P:

$$\ln P = 0.192 + 0.029t$$

Note; to properly use this expression let t = 0, in 1963 and increase it by one for each subsequent year. With this expression the value of the price index at any time in the future can be determined.

Year (1)	t (2)	Index, P (3)
1963	0	1.66
1964	1	1.71
1965	2	1.76
1966	3	1.80
1967	4	1.93
1968	5	2.09
1969	6	2.30
1970	7	2.49
1971	8	2.76
1972	9	2.95

TABLE 3.2 Price Index [Kouskoulas and Koehn 74]

4. The type variable, F, specifies the type of building. Table 3.3 provides a range of classes of buildings with their corresponding relative cost values as provided by the Department of Buildings and Safety Engineering of the City of Detroit.

Type (1)	Index, F (2)
Apartment	2.97
Hospitals	3.08
Schools	2.59
Hotels	3.08
Office building (fireproof)	2.95
Office building (not fireproof)	1.83
Stores	2.43
Garages	1.99
Factories	1.20
Foundries	1.49

TABLE 3.3 Relative Cost Index for Various Building Types [Kouskoulas and Koehn 74]

5. The height index, H, measures the height of the building by the number of stories it contains.

6. The quality variable, Q, stands for what it specifies. It is the measure of: (a) The quality of workmanship and materials used in the construction process; (b) the building use; (c) the design effort; and (d) the material type and quality used in various building components. In their article the authors let this index be equal to the average rating value of each separately ranked known building component. An arbitrary 1 to 4 scale corresponding to fair, average, good, and excellent is used to grade each component. Table 3.4 was provided to assist in the identifying and rating of building components on the basis of their qualitative description.

Component (1)	Fair (2)	Average (3)	Good (4)	Very good (5)
Use	Multitenancy	Mixed, single tenant, and Multi-tenancy	Single tenant	Single tenant with custom requirements
Design	Minimum design loads	Average design loads	Above average design loads	Many extra design loads
Exterior wall	Masonry	Glass or masonry	Glass, curtain wall, precast concrete panels	Monumental (marble)
Plumbing	Below average quality	Average quality	Above average quality	Above average quality
Flooring	Resilient, ceramics	Resilient, ceramics and terrazzo	Vinyl, ceramic terrazzo	Rug, terrazzoRw marble
Electrical	Fluorescent light, poor quality ceiling	Fluorescent light, average quality, suspended ceiling	Fluorescent light, above average quality ceiling	Fluorescent light, excellent quality ceiling
Heating, ventilating, and air conditioning	Below average quality	Average quality	Above average quality	Above average quality
Elevator	Minimum required	Above required minimum	High speed	High speed deluxe

TABLE 3.4 Quality Index [Kouskoulas and Koehn 74]

7. The technology index variable, **T**, accounts for the extra cost expended for special types of buildings or the labor and material savings resulting from the use of new techniques in the process of construction. For the usual/ordinary construction situation this variable has the value of 1. For the situation that results in extra costs **T** will be > 1 while if the employed technologies result in a cost savings the value of **T** will be $0 < T < 1$. This

variable was designed to provide the engineer/estimator with great flexibility to utilize the finally constructed cost function for the most unusual cases and furthermore to consider in his preliminary cost estimation a wide selection of technology alternatives with minimum expended time and effort. Some data regarding this variable was provided and is shown in Table 3.5.

Technology (1)	Index, T (2)
Bank-monumental work	1.75
Renovation building	0.50
Special school building	1.10
Chemistry laboratory building	1.45
Telephone building-blast resistant	1.60
County jails	1.20
Dental school	1.15
Hospital addition	1.05
Correctional center	1.20
Home for aged	1.10

TABLE 3.5 Technology Index [Kouskoulas and Koehn 74]

In summary, six variables were chosen to identify any proposed building. Of the six variables: location, year built, type, height, quality, and technology, two are very subjective, while four are quite objective. The two subjective variables, quality and technology were provided to allow the estimator latitude to utilize his/her experience to adjust the estimate for any given situation.

3.2.2 SELECTION OF THE COST-ESTIMATING RELATIONSHIP

For the form of their CER the authors arbitrarily selected a linear relationship. That is, they used techniques of multiple linear regression, employing the method of least squares, to correlate their data into the form of:

$$C = A_0 + A_1(\mathbf{L}) + A_2(\mathbf{P}) + A_3(\mathbf{F}) + A_4(\mathbf{H}) + A_5(\mathbf{Q}) + A_6(\mathbf{T}) \quad \text{.....(3.1)}$$

in which, A_k = a constant to be determined from the collected data and the **bold** letters represent the variables previously explained. The historical data used by the authors to derive their cost function is presented in Table 3.6 and the resulting cost equation is:

$$C = -81.49 + 23.93(\mathbf{L}) + 10.97(\mathbf{P}) + 6.23(\mathbf{F}) + 0.167(\mathbf{H}) + 5.26(\mathbf{Q}) + 30.9(\mathbf{T}) \quad \text{.....(3.2)}$$

Description (1)	C, in dollars per square foot (2)	V_1 (3)	V_2 (4)	V_3 (5)	V_4 (6)	V_5 (7)	V_6 (8)
Office building	36.00	0.90	2.76	2.95	40	2	1.0
Office building	25.00	0.87	2.49	2.95	18	1	1.0
Bank and office	68.50	1.02	2.76	2.95	6	4	1.75
Housing apartment	31.90	1.03	2.49	2.97	8	2	1.0
College	36.50	1.10	2.49	2.59	11	3	1.0
Renovated office building	23.30	1.13	2.95	2.95	5	2	0.5
Health science building	40.00	0.95	2.30	3.08	14	4	1.0
Telephone center	56.00	1.13	1.93	2.95	3	4	1.60
Hospital addition	40.00	1.05	2.09	3.08	5	4	1.00
Small garage	21.70	1.13	2.09	1.99	1	2	1.00
Office building	42.00	1.00	2.76	2.95	4	4	1.00
College building	45.81	1.16	2.49	2.59	1	4	1.10
Chemistry laboratory	62.00	1.16	2.95	2.59	7	4	1.45
Hospital	85.00	1.00	2.95	3.08	6	4	2.25
Dental school	47.50	1.00	2.49	3.08	7	4	1.15
Home for aged	34.30	1.13	1.93	3.08	3	3	1.10
Office building	37.00	1.13	2.76	2.95	24	1	1.00
Office building	31.90	1.13	2.30	2.95	10	2	1.00
Office building	40.00	1.13	2.30	2.95	22	3	1.00
Office building	49.50	1.13	2.95	2.95	27	3	1.00
Medical school	36.20	1.13	2.09	3.08	10	3	1.00
Union hall	24.00	1.13	2.76	1.83	1	1	1.00
Hospital addition	38.80	1.13	2.09	3.08	1	4	1.05
Office addition	20.00	1.08	1.93	2.95	4	1	1.00
College building	18.80	1.13	1.93	2.59	2	1	1.00
Office building	34.70	1.13	2.09	2.95	5	3	1.00
Office building	15.10	1.13	1.93	1.83	2	1	1.00
School, high	18.10	1.13	1.22	2.59	3	2	1.00
County correctional center	39.00	1.13	2.30	3.08	4	2	1.20
County jail	36.00	1.13	2.09	3.08	2	2	1.20
College dormitory	21.10	1.07	1.66	2.59	6	2	1.00
College dormitory	24.30	1.07	1.93	2.50	6	2	1.00
College building	30.00	1.13	1.93	2.59	6	3	1.00
Hospital addition	27.50	1.13	2.30	3.08	2	1	1.00
Foundry	11.30	1.00	2.09	1.49	1	1	1.00
Factory	14.50	1.02	2.09	1.20	1	2	1.00
Factory	10.00	1.05	2.09	1.20	1	1	1.00
Factory	14.75	0.92	2.30	1.20	1	2	1.00

TABLE 3.6 Historical Cost Data Used in Model Development

[Kouskoulas and Koehn 74])

3.2.3 MODEL TESTING

As a measure of their accuracy, the authors rely principally on the coefficient of multiple determination, R^2, as defined below:

$$R^2 = \frac{\sum_{i=1}^{n} (C_i - \underline{C})^2}{\sum_{a=1}^{n} (C_a - \underline{C})^2} \qquad \text{.....................(3.3)}$$

where:

C_a = the actual cost of the project expressed in dollars per square foot of the project (column (2) of Table 3.6).

$\underline{C}$ = the arithmetic mean of C_a.

C_i = the cost as estimated by the derived equation.

n = number of projects in sample.

From this definition it can be seen that a R^2 value of 1 would indicate that the estimated values match the actual values perfectly. In general, the closer the R^2 value is to 1 the better the fit of the regression. A R^2 value of 0 would indicate that the regression data is so scattered that no correlation or fit at all could be made. The use of the R^2 value being used as an indicator of the closeness of fit is an accepted test that is also applicable to nonlinear functions. In Kouskoulas and Koehn's paper the R^2 value is also called the measure of assumed linearity since the closer the R^2 is to 1 the closer the points are to the assumed linear plane.

For the above equation the authors obtained a R^2 value of

0.998, indicating an almost perfect correlation of C with the six variables. Additionally, the authors calculated the correlation coefficients for each variable and proved that a simpler expression with fewer variables but with an overall higher correlation results was possible by eliminating L and H. This however, according to the authors gave a poorer model in comparison to the original one since a change in the sample data towards taller buildings from a greater diversity of localities may indeed give a higher correlation value to these variables if the calculations were to be repeated. Additionally, the authors stress that in view of the fact that they are deriving a global predesign cost estimating function, the variables are necessary and essential to account for projects from different localities and involving varying building heights.

Lastly, the function was tested with an eleven story apartment-office building in Los Angeles and a thirty-nine story office building in Detroit. Quoting the authors, "the results were amazing with the difference from the actual and the estimated square foot values being only $0.10/sq ft and $0.24/sq ft respectfully.

3.2.4 ARTICLE SUMMARY

In their summary the authors stress that the true value of their work is not their actual cost function but rather the general methodology used to obtain it. Additionally, other

combinations of variables are experimented with in an effort to obtain higher R^2 values. In particular, the authors attempt to remove the subjective variables, T and Q,T respectively. Elimination of T reduces the original R^2 value from 0.998 to 0.89, while elimination of T and Q reduces it further to 0.75. Therefore, the authors conclude that subjective variables are essential and that the estimators sound judgement coupled with a thorough knowledge of the derivation are needed for the model to be accurate.

3.3 CRITIQUE OF KOUSKOULAS AND KOEHN'S WORK

In an effort to be consistent, analysis of Kouskoulas and Koehn's paper will follow the same format in which it was presented.

3.3.1 THE SELECTION OF THE INDEPENDENT VARIABLES

Of the six independent variables used by the authors, (locality, year built, type, height, quality, technology) only the height variable appears to be a true parameter. That is to say that if we define a parameter as a "cost driver", a physical characteristic of the building upon which the cost is largely dependent, (see Chapter 2), the other five variables do not conform. We have seen in Tables 3.1 through 3.5 that indexes do

exist that allow the estimator to deal with variations in location, year built, type, quality, and technology. Consequently, these factors, should not be used as variables but rather as factors that allow the user to adjust:

(1) the data base from which the cost-estimating relationship is derived so that all sample data is of the same locality, year built, type, quality and relative technology. This normalization of the data will serve to provide a basis from which other projects can effectively be estimated.

(2) a project after it has been estimated by a normalized cost-estimating function. That is to say, that the obtained value can be adjusted using indexes to correct for any particular building peculiarities that do not conform with the normalized data.

As an example of (2) above, suppose that prior to the derivation of our cost function, we adjusted the costs of our sample data projects to New York City in the year 1987. Additionally, assume that all costs were adjusted for quality, technology and type so that our data base consisted of office buildings constructed using average quality and technology. Now further suppose that we use our function to estimate a Houston, Texas apartment building in the year 1988. The building

apartment that is to be built is of average quality and technology. To estimate this project without including Kouskoulas and Koehn's five excess variables, we would first estimate the cost of the project using a derived CER (f(some independent variables) = cost). For the purposes of this example say that this value was $1,000,000. The next step is to correct for the building being located in Houston verse New York City. This is done by using the values obtained from a reputable locality index. For this example Table 3.1 will be used to obtain the following results:

$$0.9/1.16 * \$1,000,000 = \$775,862$$

From this simple computation we see that moving this apartment from a high priced area like New York City to Houston saves about $225,000. The next step in this adjustment process is to adjust our cost for the year, since our estimating equation estimates for 1987 and the project takes place in 1988. Using the Construction Cost Index published by Engineering News Record (Table 4.2 of this paper) we get:

$$1.0139 * \$775,862 = \$786,646$$

The last step in this process is to adjust for the fact the proposed building is an apartment complex as opposed to an office building. This is done by utilizing Table 3.3 to obtain:

$$2.97/1.83 * \$786,646 = \$1,276,688$$

This value of $1,276,688 is now our final estimate for an Houston apartment building. No further adjustments are necessary since the quality and technology of the proposed apartments are considered average, as is the normalized data base from which the original estimate was derived. Therefore it has been demonstrated that with the use of adjustment factors variations in projects can properly be taken into account.

As another fault of Kouskoulas and Koehn's work, one could point to the lack of a variable that relates the buildings square footage to its cost. In place of this Kouskoulas and Koehn chose to make the units of the dependent variable, cost, dollars per square foot so that their resulting estimated values are unit costs instead of a total building cost estimate. The criticism of this approach stems from the previous discussion on economies of scale in Chapter 2 and something which this model fails to consider and account. To better illustrate this point consider two buildings of different size in Table 3.7, that, when estimated utilizing the authors' approach, would render equal unit costs.

	Building 1	Building 2
Height	100 ft	100 ft
# Floors	8	8
Typical Floor Area	3000 sf	6000 sf

* Assume location, type, quality, technology and year build are identical for both buildings.

TABLE 3.7 Example of Identical Unit Costs

In the above simplistic example, the unit cost of the two buildings would be exactly the same despite of the fact that one building is twice the size of the other. This approach is considered to be quite unrealistic by this author since the presence of fixed cost that do exist in the construction industry, would automatically guarantee that the unit cost of building decrease as the size of the building increases (provided that the building costs are behaving linearly as Kouskoulas and Koehn assume). As proof of this statement the following example is provided:

#	BUILDING SF	FIXED COSTS+	VARIABLE COSTS=	TOTAL COSTS	UNIT COST
1	3000	$50,000	$500,000	$550,000	$183/SF
2	6000	$50,000	$1,000,000	$1,050,000	$175/SF

TABLE 3.8 Fictitious Building Costs

In Table 3.8 the fictitious cost of two building are compared. The first of the buildings is 3000 sf and the associated costs are as shown. The second building is twice the size of the first and as such has variable costs (assume linear relationship) twice as large as the smaller building. The fixed costs however, are by definition identical for the two buildings and as a result the larger building has more area over which to spread its fixed cost and thus has a smaller unit cost. Kouskoulas and Koehn's models' failure to account for the economies of scale is due in part to its failure to have a parameter that accounts for building dimensions other than height but also is a result of the use of a linear function to model their data. This point will be discussed in greater detail in the following section.

Another problem with Kouskoulas and Koehn's variable selection deals with the way they chose to define H, the height variable. Recall that this variable was defined to be the building height in number of stories. Use of this variable in this manner fails to account for buildings that have unusual floor heights, or for warehouse/factory type buildings that have only one story but building heights that may be 50 feet or greater. A more appropriate way to define this necessary and essential variable would be to let it represent the total height of the building in feet.

3.3.2 SELECTION OF COST-ESTIMATING RELATIONSHIP

The authors state in their article that the linear form of representing the cost-estimating relationship was arbitrarily selected. The arbitrarily selection of the cost relationship is enough reason for many to disqualify their work as a valid model based on definition alone. Strictly speaking, the formal definition of a model [McCuen 1985] says:

> " A model is simply the symbolic form in which a physical principal is expressed. It is an equation or formula but with the extremely important distinction that it was built by consideration of the pertinent physical principals, operated on by logic, and modified by experimental judgement and plain intuition. It was not simply chosen."

If this definition is used as a judging criteria, the authors' work would not qualify as a model as its linear form was simply chosen. A better approach to the selection of a functional form would have been to have fit the data from Table 3.6 to as many of the functional forms described in Chapter 2 as possible. If this had been done the authors then could have used the calculated R^2 as a basis for selection, with the best form for cost-estimating relationship being that with the R^2 value closest to 1.

3.3.3 CRITIQUE SUMMARY

In general the methods employed by Kouskoulas and Koehn to derive their cost-estimating relationship are sound. However, three critical formulation problems do exist and are as follows:

1. The height variable measures the number of stories in the building as opposed the height of the building in feet.

2. The height variable is the only one in the derived function that describes the physical dimensions of the building.

3. The form for the cost-estimating relationship was arbitrarily selected and not mathematically obtained. As a result of use of a linear form, a negative fixed costs term is included in the final CER. As mentioned in Chapter 2 the existence of negative fixed costs is an unlikely situation that normally is an indication of error.

The problems noted here were corrected in 1984 by a cost-estimating model developed by Karshenas. This model is presented and critique in the remaining sections of this chapter.

3.4 SUMMARY OF S. KARSHENAS' MODEL

3.4.1 THE SELECTION OF THE INDEPENDENT VARIABLES

Using Kouskoulas and Koehn's paper as a reference, Karshenas developed a cost-estimating relationship to estimate multistory, steel-framed office buildings [Karshenas 84]. Unlike his predecessors, Karshenas felt that only two independent variables, (height of building (ft) and the typical floor area (sf)) were necessary to adequately described the building. The other parameters that were used by Kouskoulas and Koehn in their article were considered by Karshenas but deemed unnecessary for the following reasons:

1. The type variable, T, was not needed as the author has limited his model to include only steel framed office-buildings. This approach was in fact a recommendation made by Kouskoulas and Koehn in their article as they said "...if the methodology is applied to a class of buildings instead of to the whole population of buildings, one is bound to get very good results."

2. The location variable, L, and the year variable, P, were excluded since the author instead chose to use cost and location indexes to convert all projects to March 1982, New York City cost scale. Thus, when estimating a building not

in this time period or location, adjustments based on the project specifics must be made. Using Kouskoulas and Koehn's approach these adjustments were made as part of the model.

3. Lastly, the quality variable, **Q**, and the technology variable, **T**, were omitted since the author chose only "typical buildings" in his sample. That is to say that the buildings that make up the data base do not have extraordinary floor heights or unusually wide spans. For example, an office building with a large auditorium was excluded from the data. Furthermore, as certain items were not common to all buildings, their cost was subtracted from the total cost of the buildings. Specifically, the items that were not included in the total cost are landscaping, roads, open parking spaces, waste treatment facilities, and special equipment. Thus the costs listed in column (7) of Table 3.9 represent the cost of the building itself. As a source for his cost data the author used parameter costs published by Engineering News Record.

Number (1)	Location (2)	Time of construction (3)	Number of floors* (4)	Building height, in feet (meters) (5)	Typical floor area, in square feet (square meters) (6)	Total cost, in dollars (7)	Adjusted total cost, c, in dollars* (8)
1	Lexington, Mass	Nov. 77/Jan. 79	3	36 (10.8)	27,000 (2,511)	3,133,100	4,542,000
2	Southfield, Mich.	Apr. 76/June 78	15	187.5 (56.25)	17,690 (1,645)	11,645,000	17,271,000
3	Pocatello, Idaho	May 76/Sept. 77	3	36.6 (11)	26,690 (2,502)	2,969,800	4,699,000
4	Dallas, Tex.	Apr. 76/May 77	21	262.5 (78.7)	17,000 (1,581)	12,958,000	25,254,000
5	Glendale, Calif.	Dec. 75/Nov. 76	6	72 (21.6)	15,800 (1,469.5)	2,781,480	4,506,000
6	Seattle, Wash.	Feb. 74/Dec. 76	36	468 (140.4)	22,200 (2,065)	36,470,000	70,022,000
7	Scottsdale, Ariz.	Dec. 74/May 76	2	28 (8.4)	140,332 (13,051)	12,729,900	21,050,000
8	Knoxville, Tenn.	Aug. 74/Apr. 75	2	25.3 (7.6)	9,986 (929)	446,500	928,720
9	Troy, Mich.	Aug. 73/Oct. 75	26	330 (99)	19,400 (1,804)	16,822,000	31,942,800
10	Birmingham, Ala.	Oct. 74/Jan. 76	18	216 (65)	12,616 (1,173)	6,104,140	12,704,000
11	Franklin Park, Ill.	Mar. 74/Dec. 74	5	62.6 (18.7)	8,000 (744)	1,396,200	2,619,110
12	Beverly Hills, Calif.	Nov. 73/July 75	8	105.6 (31.7)	5,500 (511.5)	1,204,100	2,175,000
13	Houston, Tex.	July 73/Jan. 75	13	175.5 (52.6)	29,920 (2,782)	10,408,000	23,722,000
14	Chicago, Ill.	Dec. 73/Dec. 74	2	28 (8.4)	35,280 (3,281)	1,951,175	3,902,350
15	Detroit, Mich.	Aug. 71/Apr. 73	4	48 (14.4)	17,700 (1,646)	1,731,800	3,300,000
16	Warren, Mich.	June 72/Oct. 73	11	137.5 (41.2)	15,000 (1,393)	4,435,000	9,275,000
17	Wellesley, Mass.	Dec. 69/Sept. 70	4	48 (14.4)	18,800 (1,748)	1,763,000	4,531,000
18	Central, N.J.	Nov. 70/Feb. 72	12	153 (45.9)	30,134 (2,802)	11,129,000	26,932,000
19	San Francisco, Calif.	Oct. 66/May 68	33	429 (128.7)	17,212 (1,600)	14,455,000	42,931,000
20	New York, N.Y.	Oct. 61/Nov. 63	42	483 (145)	18,893 (1,757)	16,820,900	63,346,000
21	Cleveland, Ohio	Feb. 63/Nov. 64	41	533 (160)	21,600 (2,009)	20,116,000	69,946,000
22	Columbus, Ohio	Dec. 63/Feb. 65	26	338 (101.4)	16,000 (1,488)	8,683,000	31,817,000
23	Pittsburgh, Pa.	Apr. 66/Apr. 68	9	126 (37.8)	16,833 (1,565)	3,871,000	11,818,000
24	Houston, Tex.	Sept. 65/Aug. 66	4	50 (15)	10,500 (977)	656,400	2,646,100

*Including basements.

TABLE 3.9 Historical Building Data [Karshenas 84]

3.4.2 SELECTION OF COST-ESTIMATING RELATIONSHIP

Unlike Kouskoulas and Koehn, who arbitrarily chose to represent their function as a linear relationship, Karshenas investigated the following types of functional forms: hyperbola, power, exponential, and logarithmic. Utilizing the graphical method described in Chapter 2, the author decided upon a power function in the form of:

$$C = z * A^{b} * H^{y}$$

for his CER. In this equation the b , y, and z are constants and

the final form of the equation, after regression analysis, using the adjusted costs in column 8 of Table 3.9, is:

$$C = A^{1.1048} * H^{1.1268} * e^{-0.0238}.$$

where, A = the typical floor area of the building (sf) and,

H = the height of the building (ft).

3.4.3 MODEL TESTING

To test whether the data was adequately described by the regression equation, Karshenas also used the coefficient of multiple determination, R^2. The R^2 value for this model was found to be 0.90 meaning that 90% of the variations in the building costs listed in Table 3.10 are accounted for by the regression equation. The remaining 10 % of the variations is due to factors not included in the model such as the quality of material and workmanship used in the building.

In the author's opinion, his power function is much more accurate then Kouskoulas and Koehn's linear cost function that expressed the square foot cost in terms of the same variables, i.e., building type, height, location, and construction year. The basis of this statement rests solely on the calculated R^2 values. In the case of Kouskoulas and Koehn, their R^2 value, when the quality and technology variables were omitted, was 0.75. On the other hand while using the power function form Karshenas was able to obtained an R^2 value of 0.90.

As a test of his model Karshenas compared the predicted square foot costs of his model with that of Means' Building System Cost Guide. The square foot costs of four proposed buildings, are shown in Table 3.10. Note that Means gives cost values for the lower quartile, median and upper quartile as shown. A comparison of the differences between the 25-percentile and the 75-percentile estimates of the two methods reveals that the proposed models' variabilities are considerably less than Means'. The author states that the interval between the 25 and 75 percentiles are less in his model due to the fact that his model estimates the cost in terms of two independent variables, the building height and the typical floor area, while Means estimates based solely on the total area of the building.

Number of stories (1)	Building height, in feet (meters) (2)	Typical floor area, in square feet (square meters) (3)	Estimated Cost,[a] in Dollars per Square Foot, Eq. 8			Estimated Cost,[b] in Dollars per Square Foot, Means		
			0.25 (4)	Median (5)	0.75 (6)	0.25 (7)	Median (8)	0.75 (9)
5	60 (18.3)	10,000 (929)	47.5	51.6	56.0	36.94	57.46	91.6
10	120 (36.5)	5,000 (464.5)	48.2	52.4	56.9	36.94	57.46	91.6
11	132 (40)	23,500 (2,181)	57.4	62.3	67.7	43.4	68.87	89.03
20	240 (73)	13,000 (1,207)	58.2	63.2	68.6	43.4	68.87	89.03

[a]Unit cost in New York City, March 1982.
[b]Means' *Building System Cost Guide*, 1982.

Table 3.10 Comparison of Predicted and Cost Book Estimates [Karshenas 84]

3.5 CRITIQUE OF KARSHENAS' PAPER

As stated previously, Karshenas, in his model corrected the principal faults of Kouskoulas and Koehn's work and as a result appears to have developed a very sound cost-estimating relationship. However, two problem areas do exist with his model and they are as follows:

1. Although Karshenas did include a parameter to account for the area of the building, he chose to let this variable, A, represent the <u>typical</u> floor area as opposed to the gross building area. As an alternative had Karshenas explored the possibility of adding the number of floors as another independent variable, or in place of the typical floor area and the number of floors variables, just have used a variable for the buildings gross floor area, better results may have been obtained.

2. When comparing his model to that of Kouskoulas and Koehn's, Karshenas claims superiority since the R^2 value of his model was 0.90 while that of his predecessors was only 0.75 when the same independent variables were used. The author attributes this to his use of a non-linear function and claims it to be the more appropriate representation of the building costs. It is the opinion of this author, that no conclusions can be drawn as to the best functional form, since different data bases were used in their derivation and

the form which best fits one set of data may not be the best fit for the next. This criticism relates back to the discussions in Chapter 2 on data collection. As stated previously, this step in the model development process is critical and inconsistencies at this stage could skew any subsequent results. In the case of the two articles, the data was collected independently from different sources and it is therefore possible that one set is more valid than the other and thus naturally gives better results.

3.6 CHAPTER SUMMARY

In this chapter two different cost-estimating models for buildings were presented. In the first article by Kouskoulas and Koehn, the authors expressed the unit cost of a building to that of six independent variables using a linear relationship. This approach was found to have its faults and was improved upon by the Karshenas' model. Karshenas' approach greatly simplified the CER by relating the cost to two independent variables through a non-linear relationship.

Together the two articles provide a solid framework for future model development. Therefore, having noted the faults of each, it is proposed that an accurate and useful cost-estimating relationship, that corrects their weaknesses, while incorporating their strengths, can be derived from historical data. The remainder of this text will be dedicated to the development, derivation and testing of this model.

CHAPTER 4 COST-ESTIMATING MODEL

4.1 INTRODUCTION

In the previous two chapters the concept of parametric estimating has been introduced and demonstrated. Attentive reading of these chapters reveals that the concepts of parametric estimating are easily understood and relatively straightforward. In fact one might conclude, as this author did, that with the use of computer software, an accurate and usable parametric model can easily be derived.

To test this hypothesis, the four steps of model development from Chapter 2 were followed and a cost-estimating relationship was derived from data obtained from three military installations in the State of Georgia.

This chapter is the description of the developed model. The format of the chapter follows that of the four steps of model development, with the content of each section being the explanation of how and why each step in the development process was handled.

4.2 PARAMETER SELECTION

As a result of the recommendations of the authors' work presented in Chapter 3, it was decided to collect cost data for only steel-framed office buildings. Using the criteria stated in

Chapter 2 as a guide, an initial list of over seventy parameters was narrowed to the following six:

1. Contract Duration
2. Amount of Liquidated Damages
3. Height of Building
4. Number of Floors
5. Typical Floor Area
6. Gross Floor Area

These parameters, as defined below, were examined/explored as possible candidates for use as cost-drivers in the final derived cost-estimating relationship.

CONTRACT DURATION (D)-is the number of days that the contractor has to complete and deliver the building. This number was thought to be significant as the inherent costs of a required accelerated construction schedule would certainly have a direct bearing on the bid price offered by any contractor.

LIQUIDATED DAMAGES (L)-is the amount of money ($/day) the contractor is accessed per day for not completing the building by the contracted completion date. As this value is an indicator of the risks being assumed by the contractor it was believed that it would be highly correlated with costs.

HEIGHT OF BUILDING (H)-is the total height measured in feet. This parameter was used by both Karshenas and Kouskoulas-Koehn in their models and will be used in the derived model as common sense dictates that the cost of a building is strongly related to its height.

NUMBER OF STORIES (S)-is the number of stories in the building. This parameter was thought to be important as it is another indicator of the size of the building.

TYPICAL FLOOR AREA (A_t)-measured in square feet, this parameter helps to further define the size of the building.

GROSS FLOOR AREA (A_g)-measured also in square feet, the value of this parameter is the result of multiplying the number of floors by the typical floor area. The use of this one variable will be explored as a substitute for the above two variables.

Although many other candidates, most of which helped to physically describe the building, were originally considered, most were eliminated because it was felt that either:

(1) the data for this particular parameters would not be available in the predesign stages of the project, or

(2) although it was a contributor to cost it was not significant enough to be used in the model. Examples of some of these parameters are:

1. Type of Roof
2. Type of Exterior Finish
3. Linear Foot of Interior Walls

4.3 DATA COLLECTION AND NORMALIZATION

4.3.1 DATA COLLECTION

Having decided upon what data was to be used the next task was the actual data collection. As a source of the building and cost information the following military installations were utilized:

1. Dobbins Air Force Base, Marietta, Georgia.
2. Fort Gillem (Army), Atlanta, Georgia.
3. Kings Bay Naval Submarine Base, Kings Bay, Georgia.

Like Karshenas, it was attempted to collect data on steel-framed office buildings only. Unfortunately, the total number of office buildings at these bases did not provide a large enough sample size to obtain significant results. As a result the search for data was expanded to include "typical" buildings from other classes. Used in this context the word "typical" is meant to mean buildings that do not contain unusual features for that particular type of building. For example a warehouse would not be excluded because it had large open areas for storage as this is typical for this class of building whereas, an office building with unusually large open areas would be excluded as being not

typical.

With the cooperation and assistance of the personnel in the construction offices of the above installations the plans and specifications for new buildings awarded in the 1980's were reviewed and considered as possible candidates. In the end, those buildings that did not contain many unusual features or specialized equipment, (that might invalidate its cost figures), were included in the sample. A summary of the data collected is contained in Table 4.1 and is self explanatory with the exception of the following:

1. The costs contained in column (7) of Table 4.1 are not in all cases the actual awarded contract prices, as the cost for unusual items, not typical for a particular building type, were subtracted from the original bid costs. For example, the cost of a large auditorium was subtracted from the cost of an office building. Additionally, the figures are for the complete project and do include landscaping, parking, overhead, profit etc.. However, no change order costs are included.

2. The costs in column (8) represent the values from column (7) after adjustments corrections for year built, location, and building type have been applied.

3. As some of the buildings had varying heights for different sections, the value in column (11) is the average building height.

Lastly, as common features, (in addition to those being made common by adjustment), do exist between the data points, any CER developed from them will be restricted for use on buildings having the same common features listed below:

a. Competitively bid buildings on military bases.

b. Steel-framed buildings.

c. Buildings with no basements.

d. Buildings with shallow foundations only.

e. Buildings not greater than three stories in height.

NUMBER (1)	PROJECT NAME (2)	CONTRACT NUMBER (3)	LOCATION (4)	YEAR BID (5)	BUILDING TYPE (6)
1	COMPOSITE SUPPORT COMPLEX	FGWB001089	DOBBINS AFB, ATLANTA,GA.	1988	OFFICE
2	ADDITION BLDG 388	FGW13000007	DOBBINS AFB, ATLANTA,GA.	1987	OFFICE
3	SQUAD. OPS BLDG	DACA21-84-C-0074	DOBBINS AFB, ATLANTA,GA.	1984	OFFICE
4	BLDG 819	DACA21-84-C-0070	DOBBINS AFB, ATLANTA,GA.	1984	OFFICE
5	WAREHOUS ADDITION	DAC21-84-C-0058	DOBBINS AFB, ATLANTA,GA.	1984	OFFICE
6	NAVAIDS FLIGHT SHOP	DAC21-85-C-0051	DOBBINS AFB, ATLANTA,GA.	1985	OFFICE
7	CONSOLIDATED AIRCRAFT SHOP	DAC21-82-C-0098	DOBBINS AFB, ATLANTA,GA.	1982	OFFICE
8	OPERATIONS FACILITY	DACA21-87-C-0073	FT. GILLEM, ATLANTA,GA.	1987	OFFICE
9	MARINE BARRACKS/ADMIN FACILITY	N68248-86-C-6018	KINGS BAY, KINGS BAY, GA.	1986	DORMITORY
10	BASE HQ/COMSUB OFFICE	N68248-84-C-4127	KINGS BAY, KINGS BAY, GA.	1987	OFFICE
11	PROVISION WAREHOUSE	N68248-84-C-4124	KINGS BAY, KINGS BAY, GA.	1987	WAREHOUSE
12	CONSOLIDATED PERSONAL SUP. CENTER	N68248-82-C-2050	KINGS BAY, KINGS BAY, GA.	1986	OFFICE
13	BACHELOR OFFICER QUARTERS	N68248-87-C-7082	KINGS BAY, KINGS BAY, GA.	1987	DORMITORY
14	SUBMARINE BASE HQ/ADMIN BLDG.	N68248-85-C-5016	KINGS BAY, KINGS BAY, GA.	1985	OFFICE
15	OFF-CREW OPERATIONS BLDG.	N68248-84-C-4125	KINGS BAY, KINGS BAY, GA.	1987	OFFICE
16	SECURITY HEADQUARTERS	N68248-84-C-4126	KINGS BAY, KINGS BAY, GA.	1987	OFFICE
17	ENLISTED QUARTERS/SUPPORT STORAGE	N68248-85-C-5029	KINGS BAY, KINGS BAY, GA.	1986	DORMITORY
18	EQUIPMENT MAINT. BLDG.	N68248-82-C-2019	KINGS BAY, KINGS BAY, GA.	1984	OFFICE
19	REFIT INDUSTRIAL BLDG.	N68248-82-C-2010	KINGS BAY, KINGS BAY, GA.	1984	OFFICE
20	PORT SERVICES SUPPORT FAC.	N68248-81-C-3044	KINGS BAY, KINGS BAY, GA.	1984	OFFICE

TABLE 4.1 DATA SUMMARY

NUMBER	AWARD PRICE ($) (7)	ADJUSTED COSTS ($) (8)	DURATION (DAYS) (9)	LIQUIDATED DAMAGES (10)	HEIGHT (FT) (11)	GROSS FLOOR AREA (SF) (12)	TYPICAL FLOOR AREA (SF) (13)	NUMBER STORIES (14)
1	$1,105,000.00	$1,105,000.00	310	110	16	15700	15700	1
2	$932,000.00	$944,954.80	340	144	16	12700	12700	1
3	$1,524,854.00	$1,641,047.87	360	390	13.6	15865	15865	1
4	$266,248.00	$286,536.10	210	122	24	3987	3987	1
5	$1,718,789.00	$1,849,760.72	390	301	20	27200	27200	1
6	$719,869.00	$765,652.67	290	201	14	9300	9300	1
7	$979,000.00	$1,142,003.50	476	250	28	21122	10561	2
8	$1,786,000.00	$1,810,825.40	350	669	20.5	11985	11985	1
9	$5,446,000.00	$5,654,563.26	420	2510	41.6	34056	8514	4
10	$1,757,680.00	$1,825,060.65	360	625	24.5	51248	25624	2
11	$2,301,000.00	$4,219,103.84	330	500	26	28914	28914	1
12	$4,230,000.00	$4,500,455.58	405	1935	19.2	40044	40044	1
13	$3,766,000.00	$3,816,129.67	450	670	28	36531	12177	3
14	$7,594,000.00	$8,271,633.58	530	900	45.75	38678	19339	2
15	$5,755,694.00	$5,976,338.47	420	1425	28	114400	57200	2
16	$2,637,665.00	$2,738,779.86	360	575	22.5	17047	17047	1
17	$7,392,000.00	$7,675,088.44	480	2300	31.8	126138	42046	3
18	$12,156,000.00	$13,397,570.32	380	650	42	24544	12272	2
19	$12,277,000.00	$13,530,928.83	500	550	37	119194	119194	1
20	$7,213,432.00	$7,950,186.12	420	500	30	30132	15066	2

TABLE 4.1 (CONT'D)

4.3.2 DATA NORMALIZATION

As described previously, normalization is the process of adjusting the data for any and all factors that differ from the norm. The norm or base for the developed model is steel-framed office buildings built in Atlanta during 1988. As a result, it was necessary to adjust data from Table 4.1 for the year built, location and building type.

The normalization of Table 4.1 was accomplished with the use of derived adjustment factors. For example to adjust for any cost differences caused by inflation due to buildings being awarded in different years, the cost was adjusted by using the Construction Cost Index published in the March 17, 1988 issue of ENR magazine. Using the indexes from this article the following adjustment factors were obtained:

TO CONVERT COST FROM YEAR	TO YEAR	MULTIPLY ORIGINAL COST BY:
1987	1988	1.0139
1986	1988	1.0389
1985	1988	1.0636
1984	1988	1.0762
1983	1988	1.0974
1982	1988	1.1665

TABLE 4.2 Cost Adjustment Factors for Year Built

To adjust for cost differences caused by varying material and labor prices that are the sole result of location, the projects from Kings Bay Georgia were adjusted, (moved to Atlanta), by using the City Cost Index from R. S. Means' Building Construction Cost Data. Applying this index resulted in the cost of the Kings Bay projects' being multiplied by 1.0241 to compensate for the cheaper material and labor prices in that part of the State.

Lastly, the buildings were adjusted for building type function by creating adjustment factors from a relative cost per square foot index for various building types. Using the data published in [Adrian 82] the following factors were developed (making office buildings the base):

TYPE OF BUILDING	MULTIPLY COST BY:
Apartments	1.2899
Banks	0.7131
Churches	1.1589
Department Stores	1.4933
Dormitories	0.9759
Factories	1.5669
Hospitals	0.6515
Libraries	1.1201
Office Buildings	1.0
Schools	1.1237
Shopping Centers	1.4172
Warehouses	1.7659

TABLE 4.3 Adjustment Factors for Building Type/Function

Using these factors each project that was not in and of itself an office type building was multiplied by the appropriate factor to adjust its costs (up or down) to account for and price variations due solely as a result of the buildings' function or type.

Additional adjustment factors that were discussed in relation to Kouskoulas and Koehn's paper, that have not yet accounted for, are the buildings quality and the technology employed during construction. In the construction of this model, no adjustments were made for these two items as it is assumed that the quality of all stateside military construction, regardless of the service branch, is roughly the same as procurement of this type is rigidly controlled and all buildings are built in compliance with the same Federal Specifications. Similarly, it has been assumed that the level of technology employed on all these projects has been roughly the same and that in general it was average.

4.4 CER SELECTION

The following four forms were investigated as possible candidates for use as the derived cost-estimating relationship:

1. Linear
2. Power
3. Exponential
4. Logarithmic

Previously these forms were explained in Chapter 2 and shown in Figure 2.1. The criteria used in selecting the best form was R-

squared values which were obtained from the multiple regression results. The software program used to perform the regression analysis was entitled Statgraphics and is marketed by the Statistical Graphics Corporation [Statgraphics 86]. This package, although fully capable of performing multiple linear regressions, was unable to adequately perform non-linear multiple regression. As a result, in order to estimate the equation for the nonlinear tested models, logarithms were used to convert the nonlinear forms to linear equations. As an example, consider the following form of the power function :

$$C = kA^{x}H^{y} \qquad (4.1)$$

in which k, x, and y are constants whose value is determined from regression analysis and A and H represent variables for area and height respectively. Taking the natural logarithms of both sides converts the original non-linear expression in equation 4.1 to the linear form expressed by equation 4.2 below:

$$\ln(C) = \ln(k) + x\ln(A) + y\ln(H) \qquad (4.2)$$

After converting the power function to its linear equivalent, the computer is capable of applying an extension of the method of least squares to perform multiple linear regression analysis and calculate the value of the coefficients. To perform such a procedure for this example, the natural logarithm would have to be first calculated for the data of each of the three variables, C, A, and H. Entering the converted values into the computer

yields the proper regression values for k, x, and y. It is important to note however, that as a result of the natural logarithm transformation, the constant value yielded for k is in reality the natural log of k making it necessary to first take the anti-natural log of the obtained value prior to using it in the final equation.

The practice of using natural logarithms to convert nonlinear forms to linear ones can easily be extended for use with exponential and logarithmic functions. Table 4.4 below summarizes the required transformation, the required input data, and the regression coefficients obtained for each of the four forms.

	Linear	Power	Exponential	Logarithmic
Equation form desired	$y = a + b_1x_1 + b_2x_2$	$y = ax_1^{b_1}x_2^{b_2}$	$y = ae^{b_1x_1}e^{b_2x_2}$	$y = a + b_1 \ln x_1 + b_2 \ln x_2$
Linear equation form	$y = a + b_1x_1 + b_2x_2$	$\ln y = \ln a + b_1 \ln x_1 + b_2 \ln x_2$	$\ln y = \ln a + b_1x_1 + b_2x_2$	$y = a + b_1 \ln x_1 + b_2 \ln x_2$
Req'd. input data transform	x_1, x_2, y	$\ln x_1, \ln x_2, \ln y$	$x_1, x_2, \ln y$	$\ln x_1, \ln x_2, y$
Regression coefficients obtained	a, b_1, b_2	$\ln a, b_1, b_2$	$\ln a, b_1, b_2$	a, b_1, b_2
Coefficient reverse transform req'd.	None	antiln $(\ln a), b_1, b_2$	antiln $(\ln a), b_1, b_2$	None
Final coefficients	a, b_1, b_2	a, b_1, b_2	a, b_1, b_2	a, b_1, b_2

*More than two independent variables are simple extensions of two variable equations.

TABLE 4.4 CER Equation Forms (Two Independent Variables)*
[Wyskida-Steward 87]

Using the adjusted costs from column (8) of Table 4.1 as the dependent variable, multiple regressions were performed, (after the appropriate linear logarithmic transformations), attempting to fit the selected independent variables to each of the four forms. For the selection process, only two sets of independent variables were used. The selected variables were chosen because they defined physical building dimensions and it was believed that they would serve as a good predictors as to how the rest of the data would behave. The adjusted R-squared values that resulted from these regressions are summarized in Table 4.5. From the results, the power function was selected as the form of the final cost-estimating relationship as its R-squared values were noticeably higher for both sets of independent variables.

ADJUSTED R-SQUARED* VALUES FOR THE DEPENDENT VARIABLE COST VS. THE INDEPENDENT VARIABLES:

EQUATION FORM	GROSS AREA	GROSS AREA & HEIGHT
Linear	0.3083	0.6693
Power	0.6120	0.7173
Exponential	0.3485	0.6178
Logarithmic	0.3619	0.6021

TABLE 4.5 Results From Initial Regression Analysis

* Adjusted R-squared values have been modified to account for the degrees of freedom. The actual R-squared values as described by equation 2.3 is slightly higher.

4.5 CER DERIVATION

Having selected the power function as the final form for the CER, the next step in the process is to decide which of the six independent variables should be used in the estimating equation.

To accomplish this tasks multiple regressions were performed on various combinations of the six independent variables. The adjusted R-squared results obtained from the analysis are shown in Table 4.6. From the R-squared values no firm conclusions could be made as to what combination of independent variables gave the best results. However, as trials 6 and 13's adjusted R-squared values were the highest they were considered to be the best candidates. A comparison of their actual R-squared values revealed that trial 6 was a slightly better fit with an actual R-squared value of 0.8239.

Given this, the residual values obtained from using the regression results from trial 6 were calculated and are shown in Table 4.7. The CER used to calculate the predicated values is:

$$C = 0.0861 * H^{1.188} * S^{-0.529} * D^{1.459} * L^{0.343} * A_g^{0.281} \quad \text{........(4.3)}$$

X INDICATES VARIABLE WAS USED IN REGRESSION

TRIAL #	ADJUSTED R-SQUARED VALUES	HEIGHT	GROSS FLOOR AREA	TYPICAL FLOOR AREA	LIQUADTED DAMADES	DURATION	NUMBER STORIES
1	0.4839	X					
2	0.6012		X				
3	0.7173	X	X				
4	0.7308	X	X			X	
5	0.7431	X	X		X	X	
6	0.7610	X	X		X	X	X
7	0.7314	X	X		X		
8	0.7080	X	X				X
9	0.7380	X	X		X		X
10	0.7080	X		X			X
11	0.7285	X		X		X	X
12	0.7380	X		X	X		X
13	0.7610	X		X	X	X	X

TABLE 4.6
R-Squared Values for Dependent Variable-Cost With the Chosen Independent Variables Marked

NUMBER	PREDICTED VALUES	RESIDUALS	%RESIDUALS
1	$757,948.84	$347,051.16	31.41%
2	$896,215.83	$48,738.97	5.16%
3	$1,203,238.02	$437,809.86	26.68%
4	$490,021.43	($203,485.33)	71.02%
5	$2,276,419.08	($426,658.36)	23.07%
6	$622,847.74	$142,804.93	18.65%
7	$2,750,138.32	($1,608,134.82)	140.82%
8	$2,091,127.25	($280,301.85)	15.48%
9	$6,411,379.82	($617,176.25)	10.65%
10	$2,742,451.79	($917,391.14)	50.27%
11	$2,949,996.06	($560,787.25)	23.47%
12	$4,836,002.17	($335,546.58)	7.46%
13	$3,344,401.14	$565,968.44	14.47%
14	$10,602,394.74	($2,330,761.16)	28.18%
15	$6,690,813.91	($714,475.44)	11.96%
16	$2,550,941.38	$187,838.48	6.86%
17	$10,179,293.29	($2,314,667.37)	29.43%
18	$4,639,665.07	$8,757,905.25	65.37%
19	$12,653,070.37	$877,858.46	6.49%
20	$3,485,399.92	$4,464,786.20	56.16%
		AVERAGE RESIDUALS	32.15%
		STANDARD DEVIATION	31.48%

TABLE 4.7 Residual Results for Equation 4.3

Review of the residual results given by Table 4.7 shows that an average error of 32.15% occurred when equation 4.3 was used to calculate the predicted costs. Further observation shows that project number 7 contained by far the highest residual percentage (140.82%). Therefore, in an effort to investigate whether better results might be obtained using these same five independent variables, a regression analysis was repeated after the data for project 7 was eliminated from the data base.

The results from this regression yielded an actual R-squared value of 0.8808 and the following CER:

$$C = 0.0016*H^{1.142}*S^{-0.345}*D^{2.606}*L^{0.170}*A^{0.125} \quad \text{..........(4.4)}$$

Using equation 4.4 to calculate the predicted cost values gives the residuals shown in Table 4.8.

The resulting higher R-squared value and the lower average percent residual value (25.74%) makes it appear that project 7 was a bad data point that introduced statistical inaccuracies into the data base. Similarly, review of the residual values contained in Table 4.8 reveals one extreme percentage (Project 14 @ 78.61%) whose removal from the data base and the subsequent regression analysis on the revised data yields an actual R-squared value of 0.9074 and the following CER coefficients:

$$C = .00001155*H^{1.457}*S^{-0.436}*D^{3.670}*L^{0.128}*A_{g}^{-0.0738} \quad \text{......(4.5)}$$

NUMBER	PREDICTED VALUES	RESIDUALS	%RESIDUALS
1	$873,144.43	$231,855.57	20.98%
2	$1,132,400.62	($187,445.82)	19.84%
3	$1,329,716.37	$311,331.50	18.97%
4	$431,228.15	($144,692.06)	50.50%
5	$2,604,594.72	($754,834.00)	40.81%
6	$653,878.63	$111,774.04	14.60%
7			
8	$2,089,498.34	($278,672.94)	15.39%
9	$6,670,866.47	($876,662.90)	15.13%
10	$2,572,846.75	($747,786.10)	40.97%
11	$2,498,236.17	($109,027.36)	4.56%
12	$3,950,545.91	$549,909.67	12.22%
13	$4,520,744.97	($610,375.40)	15.61%
14	$14,773,559.31	($6,501,925.73)	78.61%
15	$5,695,811.82	$280,526.66	4.69%
16	$2,547,008.62	$191,771.24	7.00%
17	$9,485,065.91	($1,620,439.98)	20.60%
18	$5,033,269.75	$8,364,300.57	62.43%
19	$13,389,081.72	$141,847.11	1.05%
20	$4,364,954.94	$3,585,231.18	45.10%
		AVERAGE RESIDUALS	25.74%
		STANDARD DEVIATION	20.77%

TABLE 4.8 Residual Results for Equation 4.4

The residual results for equation 4.5 are shown in Table 4.9. The value for the average residual percentage using this equation was improved to 24.27% with a noticeably lower standard deviation of 14.63%. Review of the percent residuals in Table 4.9 provides no new candidates for data elimination since no single residual percentage is clearly above the others. Additionally, further elimination of data points could severely handicap the usefulness of the model by decreasing the sample size to a statistically insignificant number. Consequently, equation 4.5 is the final form of the cost-estimating relationship derived from the data.

NUMBER	PREDICTED VALUES	RESIDUALS	%RESIDUALS
1	$809,838.64	$295,161.36	26.71%
2	$1,194,959.13	($250,004.33)	26.46%
3	$1,299,857.42	$341,190.45	20.79%
4	$392,737.55	($106,201.45)	37.06%
5	$2,842,727.84	($992,967.11)	53.68%
6	$586,111.14	$179,541.53	23.45%
7			
8	$2,331,560.21	($520,734.81)	28.76%
9	$7,649,521.51	($1,855,317.94)	32.02%
10	$2,206,936.28	($381,875.64)	20.92%
11	$2,397,969.59	($8,760.78)	0.37%
12	$3,794,558.74	$705,896.84	15.69%
13	$5,269,765.15	($1,359,395.57)	34.76%
14			
15	$4,942,961.47	$1,033,377.00	17.29%
16	$2,829,506.28	($90,726.42)	3.31%
17	$9,301,047.44	($1,436,421.52)	18.26%
18	$6,262,144.99	$7,135,425.33	53.26%
19	$16,787,750.26	($3,256,821.44)	24.07%
20	$5,273,896.47	$2,676,289.65	33.66%
		AVERAGE RESIDUALS	24.27%
		STANDARD DEVIATION	14.63%

TABLE 4.9 Residual Results for Equation 4.5

4.6 MEASURE OF GOODNESS OF FIT/MODEL TESTING

As stated previously, the R-squared values for the derivation of equation 4.5 was 0.9074. This number indicates that approximately 91% of the cost variations are accounted for by the derived model. Initially it appears that the fit for our model is pretty good. However, from the residuals in Table 4.9 it is noted that average error (24.27%) and standard deviation (14.63%), are too high, even for predesign estimating.

In an effort to put the model to a test, two projects, bid in mid August of this year, at the Kings Bay Naval Submarine were estimated using equation 4.5. The data for these projects, (which was not used to derive equation 4.5), as well as the predicated values are summarized in Table 4.10. Note that, where appropriate, the calculated cost values were adjusted for location, building type, and year. Also shown in the table are the actual high and low bids submitted by prospective contractors. For a comparison the same two projects were estimated using the median square foot costs from R. S. Means. The results along with the calculated percent residuals using both methods are also shown in Table 4.10. The table shows that the developed model performed outstandingly well in predicting the cost of the Library Building, however the results of its use in estimating the Chapel were so poor that it severely discredits the model. Its accuracy appears to be inconsistent and thus its use as an effective estimating model is doubtful.

Comparing the results obtained by equation 4.5 to those of Means makes it appear the derived model may not be total loss. However, although the Means' estimate was always at least 35% off, the main advantage it has over equation 4.5 is its consistency. This precision is necessary for any estimating method and appears to be severely lacking in the derived model.

PROJECT NAME	LIBRARY BUILDING	CHAPEL COMPLEX
PROJECT NUMBER	N68248-88-C-8052	N68248-84-C-413
LOCATION	KINGS BAY, GA.	KINGS BAY, GA.
YEAR BID	1988	1988
DURATION (DAYS)	365	480
LIQUIDATED DAMAGES ($/DAY)	325	585
HEIGHT (FT)	15.5	21.75
GROSS FLOOR AREA (SF)	16697	20960
NUMBER STORIES	1	1
PREDICTED COST USING EQUATION (4.5)	$1,775,317.73	$8,717,588.01
PREDICTED COST R. S. MEANS	$1,097,004.92	$1,402,588.70
ACTUAL BID RESULTS		
HIGH BID	$2,013,500.00	$3,169,700.00
LOW BID	$1,688,400.00	$2,587,000.00
PERCENT RESIDUALS FOR EQUATION (4.5)		
HIGH BID	11.83%	175.03%
LOW BID	5.15%	236.98%
PERCENT RESIDUALS FOR R.S MEANS		
HIGH BID	45.52%	55.75%
LOW BID	35.03%	45.78%

TABLE 4.10 Summary of Model Testing Results

CHAPTER 5 MODEL PROBLEMS

5.1 REVIEW OF THE DERIVED EQUATION'S COEFFICIENTS

In Chapter 4 a cost-estimating relationship was derived from historical cost data to test the ease of application of parametric estimating. The results of the multiple regression yielded the following power function:

$$C = 0.0000115*H^{1.457}*S^{-0.436}*D^{3.670}*L^{0.128}*A_g^{-0.0738}$$

Use of this equation to estimate two buildings at the Kings Bay Naval Submarine Base provided mixed results indicating that the model may be flawed. A closer look at the individual coefficients confirms this belief as it appears that the some of the selected parameters are not correlated to the total cost in the manner that was originally intended when they were selected. Looking at the coefficients for each variable individually reveals that:

1. Both the height and the liquidated damages variables appear to be functioning properly as they both have positive coefficients that causes the cost of the building to go up when either the height or the amount charged for damages are increased.

2. The coefficient for the number of stories is negative. As a result, as the number of stories increases the value of this variable becomes a smaller and smaller fraction thus reducing the total cost. At first thought this appears to be incorrect as the cost of the building should be increasing as the number of stories are increased, however this variable may be one of the ways in which the principle of the economies of scale has manifested itself in the model.

3. A look at the coefficient for duration reveals the largest positive value in the model. Consequently, as the number of days are increased the cost of the project is also increased. Originally, this parameter was considered for use in the model to account for the increased costs resulting from the contractor not being allowed adequate time to complete the project. It was intended for this parameter to be negatively correlated with cost so that as the number of days to complete the work decreased the cost of the contract increased. In actuality, the regression analysis saw that the projects that had longer durations cost more and thus related the two directly. Since this parameter is not being used in the derived model as originally defined, its inclusion as a valid parameter is suspect.

4. Lastly, the coefficient for the gross floor area is incorrectly a small negative number. This becomes clear by simply assigning fixed values to all variables except the gross area, which is increased. The values computed for this component of the equation decrease as the square footage increases, thus making the total cost of the project decrease as the gross square footage is increased.

As a result of 3 and 4 above, it must be concluded that the model does not perform in a realistic manner that reflects the true nature of building cost, therefore its use as a valid means of estimating is not recommended.

5.2 SOURCE OF ERROR

In Chapter 3 we saw, the concepts of parametric estimating could be successfully applied to data gathered from published sources to develop a relationship to estimate buildings. However, when applied to actual data collected from the field this technique provided erratic results. As a consequence, it is believed that the principle source of the problems incurred by the model derived in Chapter 4 occurred at the data collection stage of the process. In general the exactness of data collection required by parametric estimating is its chief limitation. More specifically, with regards to the developed model, the most demanding step in its development was the data

selection and subsequent collection. Great care was taken in reviewing the plans and specifications of all data point candidates so that the cost for non-common items could be eliminated. In retrospect it appears that the task of reviewing hundreds of sheets of drawings and thousands of pages of specifications for projects, with the hopes of uncovering the majority of non-common features, was unrealistic. Possibly if the reviewer had been involved in the construction or design of the buildings, a task of this kind could successfully be undertaken. Not surprisingly the source of the data for both articles from Chapter 3 came from published sources and thus eliminated the need for the authors to review plans and specifications.

CHAPTER 6

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS FOR FUTURE RESEARCH

6.1 SUMMARY

This paper has introduced parametric estimating as a fast, inexpensive, reasonable accurate, alternative method of estimating the cost of a building before the detailed plans and specifications are available. The methodology for the development of a parametric model was found to be:

1. Parameter Selection
2. Data Collection and Normalization
3. CER Form Selection and Derivation
4. Measuring the Goodness of Fit/Model Testing

The use and actual application of these four steps was illustrated by the presentation and critique of two previously published cost-estimating models.

In an effort to test the ease of application of this estimating technique, a parametric estimating model was developed from cost data collected from three Georgia military installations. The resulting cost estimating function related the total cost of a building to the following five independent variables: gross floor area, number of stories, length of contract, liquidated damages, and height. Test of this derived

estimating relationship indicted that problems with its formulation did exist and as result the model was not suitable for use.

6.2 CONCLUSIONS

In theory, the concepts behind parametric estimating are straightforward and relatively easily understood thus making it appear that the development of an accurate and effective estimating model is easily accomplished. In principle this is true, however, in actuality this technique provides many opportunities for error that make it quite difficult for a usable model to be developed. In the case of the model developed in Chapter 4, a data base containing both inaccurate cost data and too few data points, as well as poor parameter selection was the apparent cause of the model's failure.

Although the two models presented in Chapter 3 were derived from published data it is believed that this technique can still be accurately applied to actual field data if the steps in the development process are followed properly. However, to successfully accomplish this it is felt that a person, who is not only knowledgeable of parametric estimating and the statistics behind it, but also one who is intimately familiar with the projects being considered and selected for entry in the data base, is needed to undertake the effort. This is thought to be true as only a person with this background has the necessary knowledge to:

1. Properly select parameters that are known to have a high correlation with cost for the particular type projects being reviewed.

2. Effectively and consistently exclude projects from the data base due to abundance of non-common features or specialized equipment.

3. Properly subtract individual costs, (from the total project cost), for those atypical items that are not severe enough to have the data excluded from the data base.

In summary, the method of parametric estimating is not as simplistic as it appears from the surface. As such its use appears to be limited to the experienced estimator, who is both knowledgeable of parametric techniques and the projects in the data base. Use of this technique by others to create a model providing predesign building estimates will probably yield poor results.

6.3 RECOMMENDATIONS FOR FUTURE RESEARCH

Of apparent interest to this field would be a study attempting to identify where in the construction industry parametric estimating is being successfully employed. Included in this study would be the finding of what was being estimated,

the methods of application, and the accuracies being obtained. From these findings the development of subsequent parametric models would be greatly simplified with much better results being attainable.

REFERENCES

[Adrian 82] Adrian, J., Construction Estimating, Reston Publishing Company Inc., 1982

[Calder 76] Calder G., The Principles and Techniques of Engineering Estimating, Pergamon Press, 1976

[Dean 76] Dean, J., Statistical Cost Estimation, Indiana University Press, 1976

[ENR 88] "First Quarterly Cost Report", Engineering News Record, March 17, 1988

[Gallagher 82] Gallagher P., Parametric Estimating for Executives and Estimators, Van Nostrand Reinhild Company, 1982

[Halpin-Woodhead 80] Halpin D., and Woodhead, R., Construction Management, John Wiley & Sons, 1980

[Hines-Montgomery 80] Hines, W. and Montgomery D., Probability and Statistics in Engineering and Management Science, John Wiley & Sons Inc, 1980

[Karshenas 84] Karshenas, S.,"Predesign Cost-Estimating Method for Multistory Buildings", ASCE Journal of the Construction Engineering and Management, 1984

[Kouskoulas-Koehn 75] Kouskoulas, V. and Koehn, E., "Predesign Cost-Estimating Function for Buildings", ASCE Journal of the Construction Division, September 1975

[McCuen 85] McCuen R., Statistical Methods For Engineers, Prentice Hall, Inc., 1985

[Means 86] Means Square Foot Costs 1987, R. S. Means Company, Inc., 1986

[Ostwald 84] Ostwald P., Cost Estimating, Prentice Hall, Inc.,1985

[Statgraphics 86] Statistical Graphics Corporation, 1986

[Wyskida-Steward 87] Wyskida, R. and Steward, R., Cost Estimator's Reference Manual, John Wiley & Sons Inc., 1987

APPENDIX A

REGRESSION RESULTS FOR TABLE 4.5

Model fitting results for: cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	1.89491E6	1.10118E6	1.7208	0.1024
grossarea	63.941855	20.781781	3.0768	0.0065

R-SQ. (ADJ.) = 0.3083 SE= 3339904.557930 MAE= 2374672.021137 DurbWat= 1.502
Previously: 0.6177 0.638822 0.415485 2.259
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	-2.523631E7	8.661035E6	-2.9138	0.0093
LOG grossarea	2.900656E6	8.452223E5	3.4318	0.0030

R-SQ. (ADJ.) = 0.3619 SE= 3207707.543219 MAE= 2254962.225363 DurbWat= 1.358
Previously: 0.3083 3339904.557930 2374672.021137 1.502
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	-3.178494E7	7.093808E6	-4.4807	0.0003
LOG grossarea	1.481162E6	7.83732E5	1.8899	0.0759
LOG height	6.548626E6	1.898232E6	3.4499	0.0031

R-SQ. (ADJ.) = 0.6026 SE= 2531458.727405 MAE= 1825708.782268 DurbWat= 1.253
Previously: 0.7173 0.549373 0.388413 2.343
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	12.772223	0.433986	29.4300	0.0000
grossarea	0.000011	4.330651E-6	2.5354	0.0213
height	0.062483	0.016894	3.6985	0.0018

R-SQ. (ADJ.) = 0.6177 SE= 0.638822 MAE= 0.415485 DurbWat= 2.259
Previously: 0.6026 2531458.727405 1825708.782268 1.253
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	5.291483	1.761701	3.0036	0.0076
LOG grossarea	0.936093	0.171923	5.4448	0.0000

```
R-SQ. (ADJ.) = 0.6012  SE=      0.652465  MAE=       0.494621  DurbWat=  2.032
Previously:    0.3485           0.833996             0.627420            1.799
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.
```

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	14.175622	0.274972	51.5530	0.0000
grossarea	0.000017	5.189346E-6	3.3409	0.0036

```
R-SQ. (ADJ.) = 0.3485  SE=      0.833996  MAE=       0.627420  DurbWat=  1.799
Previously:    0.6012           0.652465             0.494621            2.032
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.
```

Model fitting results for: cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	-4.337302E6	1.569211E6	-2.7640	0.0133
grossarea	35.71032	15.658821	2.2805	0.0357
height	2.774729E5	6.108554E4	4.5424	0.0003

R-SQ. (ADJ.) = 0.6691 SE= 2309858.492943 MAE= 1712919.982550 DurbWat= 1.372
Previously: 0.3485 0.833996 0.627420 1.799
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	4.098287	1.539485	2.6621	0.0164
LOG grossarea	0.677454	0.170084	3.9831	0.0010
LOG height	1.193194	0.411951	2.8964	0.0100

R-SQ. (ADJ.) = 0.7173 SE= 0.549373 MAE= 0.388413 DurbWat= 2.343
Previously: 0.6691 2309858.492943 1712919.982550 1.372
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

APPENDIX B

REGRESSION RESULTS FOR TABLE 4.6

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	8.248058	1.531454	5.3858	0.0000
LOG height	2.054638	0.473755	4.3369	0.0004

```
R-SQ. (ADJ.) = 0.4838  SE=        0.742328  MAE=        0.490956  DurbWat=  2.013
Previously:    0.6177             0.638822              0.415485             2.259
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.
```

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	5.291483	1.761701	3.0036	0.0076
LOG grossarea	0.936093	0.171923	5.4448	0.0000

```
R-SQ. (ADJ.) = 0.6012  SE=        0.652465  MAE=        0.494621  DurbWat=  2.032
Previously:    0.4838             0.742328              0.490956             2.013
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.
```

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	-2.629389	4.419562	-0.5949	0.5607
LOG floorarea	0.469548	0.232104	2.0230	0.0613
LOG height	1.258547	0.498316	2.5256	0.0233
LOG duration	1.476282	0.994645	1.4842	0.1585
LOG stories	0.125657	0.427073	0.2942	0.7726

R-SQ. (ADJ.) = 0.7285 SE= 0.538414 MAE= 0.349413 DurbWat= 2.374
Previously: 0.7080 0.558287 0.390893 2.385
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	3.416948	1.850099	1.8469	0.0846
LOG floorarea	0.510061	0.20313	2.5110	0.0240
LOG height	1.315547	0.483769	2.7194	0.0158
LOG damages	0.346612	0.206078	1.6819	0.1133
LOG stories	0.065366	0.426913	0.1531	0.8804

R-SQ. (ADJ.) = 0.7380 SE= 0.528877 MAE= 0.348011 DurbWat= 1.824
Previously: 0.7285 0.538414 0.349413 2.374
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	3.416948	1.850099	1.8469	0.0846
LOG grossarea	0.510061	0.20313	2.5110	0.0240
LOG height	1.315547	0.483769	2.7194	0.0158
LOG damages	0.346612	0.206078	1.6819	0.1133
LOG stories	-0.444695	0.37557	-1.1841	0.2548

R-SQ. (ADJ.) = 0.7380 SE= 0.528877 MAE= 0.348011 DurbWat= 1.824
Previously: 0.7080 0.558287 0.390893 2.385
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	3.30561	1.951729	1.6937	0.1097
LOG floorarea	0.703099	0.176919	3.9741	0.0011
LOG height	1.389347	0.508565	2.7319	0.0148
LOG stories	0.445931	0.382152	1.1669	0.2604

R-SQ. (ADJ.) = 0.7080 SE= 0.558287 MAE= 0.390893 DurbWat= 2.385
Previously: 0.7380 0.528877 0.348011 1.824
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	4.098287	1.539485	2.6621	0.0164
LOG grossarea	0.677454	0.170084	3.9831	0.0010
LOG height	1.193194	0.411951	2.8964	0.0100

R-SQ. (ADJ.) = 0.7173 SE= 0.549373 MAE= 0.388413 DurbWat= 2.343
Previously: 0.6012 0.652465 0.494621 2.032
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	3.30561	1.951729	1.6937	0.1097
LOG floorarea	0.703099	0.176919	3.9741	0.0011
LOG height	1.389347	0.508565	2.7319	0.0148
LOG stories	0.445931	0.382152	1.1669	0.2604

R-SQ. (ADJ.) = 0.7080 SE= 0.558287 MAE= 0.390893 DurbWat= 2.385
Previously: 0.7173 0.549373 0.388413 2.343
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	4.643577	1.551994	2.9920	0.0086
LOG grossarea	0.509965	0.205666	2.4796	0.0247
LOG height	1.021675	0.420434	2.4300	0.0272
LOG damages	0.274175	0.199245	1.3761	0.1878

R-SQ. (ADJ.) = 0.7314 SE= 0.535480 MAE= 0.377196 DurbWat= 1.851
Previously: 0.7610 0.505154 0.310643 1.795
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	3.30561	1.951729	1.6937	0.1097
LOG grossarea	0.703099	0.176919	3.9741	0.0011
LOG height	1.389347	0.508565	2.7319	0.0148
LOG stories	-0.257168	0.378583	-0.6793	0.5067

R-SQ. (ADJ.) = 0.7080 SE= 0.558287 MAE= 0.390893 DurbWat= 2.385
Previously: 0.7314 0.535480 0.377196 1.851
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	-2.447128	4.147869	-0.5900	0.5646
LOG grossarea	0.281229	0.243078	1.1569	0.2667
LOG height	1.187055	0.469328	2.5293	0.0241
LOG duration	1.45837	0.933259	1.5627	0.1404
LOG damages	0.343226	0.196846	1.7436	0.1031
LOG stories	-0.528533	0.362714	-1.4572	0.1671

R-SQ. (ADJ.) = 0.7610 SE= 0.505154 MAE= 0.310643 DurbWat= 1.795
Previously: 0.7431 0.523728 0.337207 1.751
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	-2.447128	4.147869	-0.5900	0.5646
LOG floorarea	0.281229	0.243078	1.1569	0.2667
LOG height	1.187055	0.469328	2.5293	0.0241
LOG duration	1.45837	0.933259	1.5627	0.1404
LOG damages	0.343226	0.196846	1.7436	0.1031
LOG stories	-0.247304	0.45421	-0.5445	0.5947

R-SQ. (ADJ.) = 0.7610 SE= 0.505154 MAE= 0.310643 DurbWat= 1.795
Previously: 0.7610 0.505154 0.310643 1.795
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	-1.008183	4.044117	-0.2493	0.8063
LOG grossarea	0.459237	0.230859	1.9893	0.0641
LOG height	1.015749	0.422668	2.4032	0.0287
LOG duration	1.330099	0.977996	1.3600	0.1927

R-SQ. (ADJ.) = 0.7307 SE= 0.536138 MAE= 0.353695 DurbWat= 2.262
Previously: 0.7080 0.558287 0.390893 2.385
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	-0.212314	3.995615	-0.0531	0.9583
LOG grossarea	0.31268	0.25102	1.2456	0.2320
LOG height	0.863144	0.428546	2.0141	0.0623
LOG duration	1.257218	0.95693	1.3138	0.2087
LOG damages	0.259483	0.195193	1.3294	0.2036

R-SQ. (ADJ.) = 0.7431 SE= 0.523728 MAE= 0.337207 DurbWat= 1.751
Previously: 0.7307 0.536138 0.353695 2.262
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	-2.447128	4.147869	-0.5900	0.5646
LOG grossarea	0.281229	0.243078	1.1569	0.2667
LOG height	1.187055	0.469328	2.5293	0.0241
LOG damages	0.343226	0.196846	1.7436	0.1031
LOG stories	-0.528533	0.362714	-1.4572	0.1671
LOG duration	1.45837	0.933259	1.5627	0.1404

R-SQ. (ADJ.) = 0.7610 SE= 0.505154 MAE= 0.310643 DurbWat= 1.795
Previously: 0.7380 0.528877 0.348011 1.824
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Analysis of Variance for the Full Regression

Source	Sum of Squares	DF	Mean Square	F-Ratio	P-value
Model	16.7110	5	3.34220	13.0974	.0001
Error	3.57253	14	0.255181		
Total (Corr.)	20.2835	19			

R-squared = 0.82387
R-squared (Adj. for d.f.) = 0.760967

Stnd. error of est. = 0.505154
Durbin-Watson statistic = 1.79532

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	-2.447128	4.147869	-0.5900	0.5646
LOG floorarea	0.281229	0.243078	1.1569	0.2667
LOG height	1.187055	0.469328	2.5293	0.0241
LOG damages	0.343226	0.196846	1.7436	0.1031
LOG stories	-0.247304	0.45421	-0.5445	0.5947
LOG duration	1.45837	0.933259	1.5627	0.1404

R-SQ. (ADJ.) = 0.7610 SE= 0.505154 MAE= 0.310643 DurbWat= 1.795
Previously: 0.7610 0.505154 0.310643 1.795
20 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Analysis of Variance for the Full Regression

Source	Sum of Squares	DF	Mean Square	F-Ratio	P-value
Model	16.7110	5	3.34220	13.0974	.0001
Error	3.57253	14	0.255181		
Total (Corr.)	20.2835	19			

R-squared = 0.82387 Stnd. error of est. = 0.505154
R-squared (Adj. for d.f.) = 0.760967 Durbin-Watson statistic = 1.79532

APPENDIX C

REGRESSION RESULTS FOR EQUATIONS 4.4 AND 4.5

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	-11.368987	4.103352	-2.7707	0.0169
LOG height	1.456829	0.380506	3.8287	0.0024
LOG stories	-0.435654	0.280232	-1.5546	0.1460
LOG duration	3.668582	0.941766	3.8954	0.0021
LOG damages	0.128104	0.158814	0.8066	0.4356
LOG grossarea	-0.073767	0.210801	-0.3499	0.7325

R-SQ. (ADJ.) = 0.8688 SE= 0.375837 MAE= 0.260125 DurbWat= 2.499
Previously: 0.0000 0.000000 0.000000 0.000
18 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Analysis of Variance for the Full Regression

Source	Sum of Squares	DF	Mean Square	F-Ratio	P-value
Model	16.6112	5	3.32223	23.5196	.0000
Error	1.69504	12	0.141254		
Total (Corr.)	18.3062	17			

R-squared = 0.907406 Stnd. error of est. = 0.375837
R-squared (Adj. for d.f.) = 0.868825 Durbin-Watson statistic = 2.49927

Model fitting results for: LOG cost

Independent variable	coefficient	std. error	t-value	sig.level
CONSTANT	-6.435863	3.77868	-1.7032	0.1123
LOG height	1.141143	0.392649	2.9063	0.0123
LOG stories	-0.344412	0.310998	-1.1074	0.2882
LOG duration	2.604581	0.891645	2.9211	0.0119
LOG damages	0.17013	0.176986	0.9613	0.3540
LOG grossarea	0.125289	0.211495	0.5924	0.5637

R-SQ. (ADJ.) = 0.8348 SE= 0.422211 MAE= 0.257693 DurbWat= 1.906
Previously: 0.0000 0.000000 0.000000 0.000
19 observations fitted, forecast(s) computed for 0 missing val. of dep. var.

Analysis of Variance for the Full Regression

Source	Sum of Squares	DF	Mean Square	F-Ratio	P-value
Model	17.1089	5	3.42178	19.1952	.0000
Error	2.31740	13	0.178262		
Total (Corr.)	19.4263	18			

R-squared = 0.880708 Stnd. error of est. = 0.422211
R-squared (Adj. for d.f.) = 0.834826 Durbin-Watson statistic = 1.90562

Thesis
M188243 McGarrity
c.1 Parametric estimating.