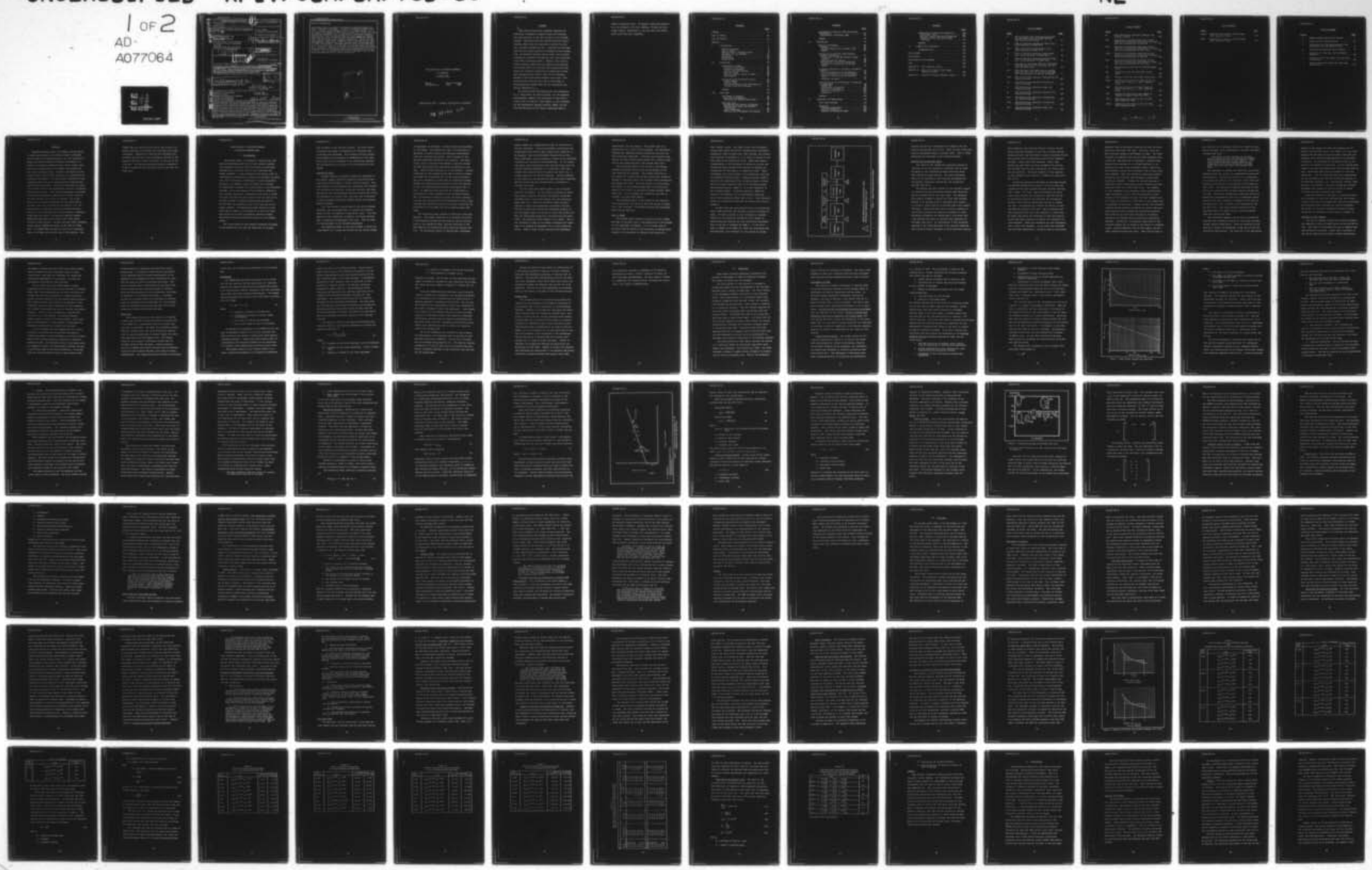
AD-A077 064 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOO--ETC F/G 15/5
APPLICATION OF A BAYESIAN APPROACH TO UPDATING AIRFRAME CERS.(U)
SEP 79 W D DIETRICH
UNCLASSIFIED AFIT/GSM/SM/76D-30 NL

1 OF 2
AD-
A077064



AD A077064

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: GSM/SM/76D-30	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Application of a Bayesian Approach to Updating Airframe CERs.	5. TYPE OF REPORT & PERIOD COVERED: M.S. Thesis
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Walter D. Dietrich, Captain, USAF	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Air Force Institute of Technology (AFIT/EN) Wright-Patterson AFB, Ohio 45433	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE: Sept[redacted]979
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

(6) (10) (9) Master's thesis (11) (12) 179

LEVEL II

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited. ~~Appendices A, B, and C are limited to U.S. Government agencies only; proprietary information, September 1979.~~ ~~Other requests for Appendices (A, B, and C) of this document must be referred to Dean of Engineering, Air Force Institute of Technology (AFIT/EN), Wright-Patterson Air Force Base, Ohio 45433.~~

B041775

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(14) AFIT/GSM/SM/76D-30

DDC RECEIVED NOV 7 1979 E

18. SUPPLEMENTARY NOTES

Approved for public release; IAW 190-17

JOSEPH P. HIPPS, Major, USAF
Director of Information

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Parametric Cost Estimation
Aircraft Airframe
Parametric Airframe Cost Estimation
Airframe Cost Estimation
Aircraft Cost Estimation
Airframe CERs
Bayesian Approach

012 225 Gw

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Because production costs of airframe programs absorb a significant portion of the defense budget, government decision makers and program managers use cost estimates to assess and decide follow-on procurement strategies, to budget for follow-on production buys, and to control obligations and expenditures of the budget appropriated for the current production buy. Production estimates are performed using a "grass roots" approach or an application of learning-curve theory. The "grass roots" approach is time consuming, while the use of learning-curves requires subjectivity by the analyst. Parametric techniques are quick

DDC FILE COPY

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 (Continuation)

and eliminate much of the subjectivity inherent in applying learning-curve theory. This study investigates an approach for developing parametric equations to estimate costs of the next lot/unit of an aiframe program. Due to the nature of the problem, estimating the recurring cost of the next unit using all available information (i.e. lot observations of historical airframe programs and early actual lot costs available for the new program) a Bayesian methodology was chosen. The Bayesian approach updates RANDOM CERs developed from a mixed linear model which considers random effects. Because the basic model considers random affects, RANDOM was chosen as the name for these CERs. The RANDOM CERs consider an error due to different types of airframes and an error due to the equation. The RANDOM CERs were chosen because these CERs proved to be better predictors of airframe cost than CERs using other techniques. Comparison of Bayesian CER estimates with estimates derived using other techniques resulted in the Bayesian CER being a better predictor of the cost of the next unit. The Bayesian approach provides a parametric means of estimating the functional costs of the next airframe unit.

APPLICATION OF A BAYESIAN APPROACH
TO UPDATING
AIRFRAME CERs

Thesis
GSM/SM/76D-30

Walter Dietrich
Capt USAF

Preface

This study investigates a Bayesian approach for developing a parametric equation which will estimate the recurring cost of the next lot/unit of an airframe program. Recurring costs are predicted because definitionally these costs are expected to reflect the cost for a follow-on production unit. Although the data base used for this study consisted of production cost information, the Bayesian approach may be useful for providing a parametric estimate of production cost using recurring costs from a prototype effort. However, until definitional problems associated with separating engineering and tooling costs into recurring and nonrecurring categories are resolved, predictions of production or next unit engineering and tooling costs will be marginal. Because of the definitional problem, total cost, (nonrecurring and recurring) was used in this study to develop Bayesian updated CERs for the engineering and tooling categories.

My sincere thanks and appreciation are extended to Dr. N. Keith Womer, my thesis advisor, for his patience, understanding, support, and assistance in this effort. I would like to thank Mr. Chuck Samson, a cost estimator for the Aeronautical Systems Division, WPAFB, for his time and discusion of the current techniques used to predict production costs. My deepest thanks and appreciation are extended to my wife, Barbara. Without her continued support, understanding, and sacrifice this effort would never have been completed.

Contents

Contents

Contents

List of Tables

List of Tables

List of Tables

List of Figures

Abstract

Because production costs of airframe programs absorb a significant portion of the defense budget, government decision makers and program managers use cost estimates to assess and decide follow-on procurement strategies, to budget for follow-on production buys, and to control obligations and expenditures of the budget appropriated for the current production buy. Production estimates are performed using a "grass roots" approach or an application of learning-curve theory. The "grass roots" approach is time consuming, while the use of learning-curves requires subjectivity by the analyst. Parametric techniques are quick and eliminate much of the subjectivity inherent in applying learning-curve theory. This study investigates an approach for developing parametric equations to estimate costs of the next lot/unit of an airframe program. Due to the nature of the problem, estimating the recurring cost of the next unit using all available information (i.e. lot observations of historical airframe programs and early actual lot costs available for the new program) a Bayesian methodology was chosen. The Bayesian approach updates RANDOM CERs developed from a mixed linear model which considers random effects. Because the basic model considers random affects, RANDOM was chosen as the name for these CERs. The RANDOM CERs consider an error due to different types of airframes and an error due to the equation. The

RANDOM CERs were chosen because these CERs proved to be better predictors of airframe cost than CERs using other techniques. Comparison of Bayesian CER estimates with estimates derived using other techniques resulted in the Bayesian CER being a better predictor of the cost of the next unit. The Bayesian approach provides a parametric means of estimating the functional costs of the next airframe unit.

APPLICATION OF A BAYESIAN APPROACH TO UPDATING AIRFRAME CERs

I. Introduction

This thesis effort is conducted to develop cost estimating relationships that will provide accurate updated cost estimates of future lots of airframes as additional information of the actual cost of prior produced lots become available. A cost estimating relationship (CER) is a mathematical formula which predicts the dependent variable, cost, from a functional relationship of known independent variables, for example, weight and speed for airframes. The independent variables are causal effects of the dependent variable. Due to the nature of the problem of updating a cost estimate using prior information (historical airframe cost data) and observed data (early actual lot costs of an airframe program), a Bayesian approach was selected to develop CERs for estimating costs of the next lot of airframes produced. The Bayesian approach employed in this thesis may be useful for providing lot or unit cost estimates which can be used to determine production budget requirements during the production phase of a new weapons system.

The first section of this chapter discusses the phases of the acquisition cycle and the importance of accurate cost estimates in the decision process. The next section introduces the reader to the spectrum of techniques used to provide cost estimates. The following sections present the statement of the problem, an enumeration of the objectives, and a brief discussion of the methodology employed. The last section is an outline of the organization of the thesis.

Acquisition Cycle

Systems analysis has become increasingly important in this modern complex world for evaluating alternatives and aiding the decision maker in making informed choices. Systems analysis is a methodology that assimilates the theoretical principles of economics, operations research, and other disciplines into the managerial decision process (Ref 31:12). The Department of Defense (DOD) uses this type of analysis to select between proposed new weapon systems and existing weapon system(s).

For a new system to become part of the active defense inventory, the new program must go through five distinct phases of what is termed the "acquisition cycle". The five phases are the conceptual, validation, full-scale development, production and deployment (Ref 4:43). Each of these will be discussed in the following paragraphs.

The conceptual phase is the initial phase in which the requirements for a system are defined and the system concept of employment is developed. Several studies are performed in this phase. The feasibility study is accomplished to determine if current design and production technologies exist for producing the system. Then estimates of the desired operational suitability are prepared. The feasibility study is concerned with whether a specific design can meet the operational requirements established by the user. Another study performed is a cost analysis. Current techniques of cost estimation are used to provide information about the preliminary cost of the proposed program. This cost information is then used for comparison studies between the proposed program and the various alternatives to meet the established operational requirements (Ref 3: 2-4). An important study that must be accomplished in this phase is a risk assessment of the proposed program. This is performed to explicitly state the known uncertainties and risk involved with the program which allows a more informed and intelligent decision to be made concerning the transition of the program into the validation phase (Ref 4:48).

The validation phase consists of additional study and analysis. The primary objective is to validate the choice of the proposed program selected from the various alternatives in the conceptual phase, and then determine if the next phase of the acquisition cycle should be entered (Ref 2:4). The preliminary design is defined when independent design studies are accomplished by both the contractor(s) and the government. Definite performance requirements are established which in many cases require cost trade-offs to be analyzed. Now that performance characteristics have been established, various parametric studies can be performed including cost estimations and evaluations. In the validation phase, some experimental shop fabrication may occur on various subsystem components or a prototype, fully functioning total system, may be required to be built (e.g. a flyable system for an aircraft program). The key elements of this phase are the experimental shop fabrication approach and the fly-off and test concepts when a prototype is required (Ref 4:50).

The full-scale development phase is the transition phase for the system. In this phase the system and necessary support items are designed, fabricated, tested and evaluated prior to the system moving into the production phase (Ref 2:4). While the validation phase required testing to ensure the performance requirements are met, testing of additional systems and subsystems produced in the development phase is required to ensure the system will meet all established operational configurations and objectives. Two important events happen in this phase. One is the initiation of many of the formal procedures and techniques to be employed by management for a large production effort. Some of these include configuration management, contracting, and cost control. The second event is a specification of a final detailed design. The engineering drawings are completed so that contracting can occur for force structure quantities. Efficient and successful completion of tasks in this phase provide for a smooth transition to the production phase that follows (Ref 4:60).

The production phase is the period of time between production approval and delivery and acceptance of the last system by the customer. The primary objective of this phase is to efficiently produce and deliver effective systems to operating units, and to ensure these systems are supportable (Ref 2:4). Once the system is in this phase, any changes in the system hardware must be formally presented in the form of an engineering change proposal (ECP), (Ref 4:74). As the systems are accepted by the user or customer, the next phase begins.

The deployment phase is the period of time from the first operational unit delivery until the system is phased out of the inventory. The production phase and deployment phase overlap (Ref 2:4).

Role of DSARC

The Defense System Acquisition Review Council (DSARC) performs an important role in the acquisition of new systems for the Department of Defense. It is a formal body of officials from the Office of the Secretary of Defense which reports to the Secretary of Defense on the status of a major weapons system. The DSARC reviews the Development Concept paper which defines the program plans, performance parameters, areas of major risk, cost analyses, and acquisition strategy to determine if it is ready to proceed to the next phase of the acquisition cycle. DSARC then makes a recommendation to the Secretary of Defense for action concerning the major program which consists of approval or disapproval for continuance, or additional study. These proceedings occur three times in the acquisition cycle and are important decision points. Program decision, DSARC I, occurs at the completion of the conceptual phase to determine if progress should be made to the validation phase. Ratification decision, DSARC II, is at the completion of the validation phase. Production decision, DSARC III, occurs at the completion of the development phase and prior to the production phase (Ref 3:a1-1). These decision points are illustrated along with the acquisition cycle in figure 1.

To make decisions intelligently during the acquisition cycle, DSARC performs cost analyses to obtain estimates of costs. The basic purpose of cost estimating in systems analysis is to provide an indication of the amount of scarce resources required for each alternative considered. This is the use of estimates by DSARC I. As the cycle moves to DSARC II and DSARC III, where the alternative has been selected, cost estimates for the program or system

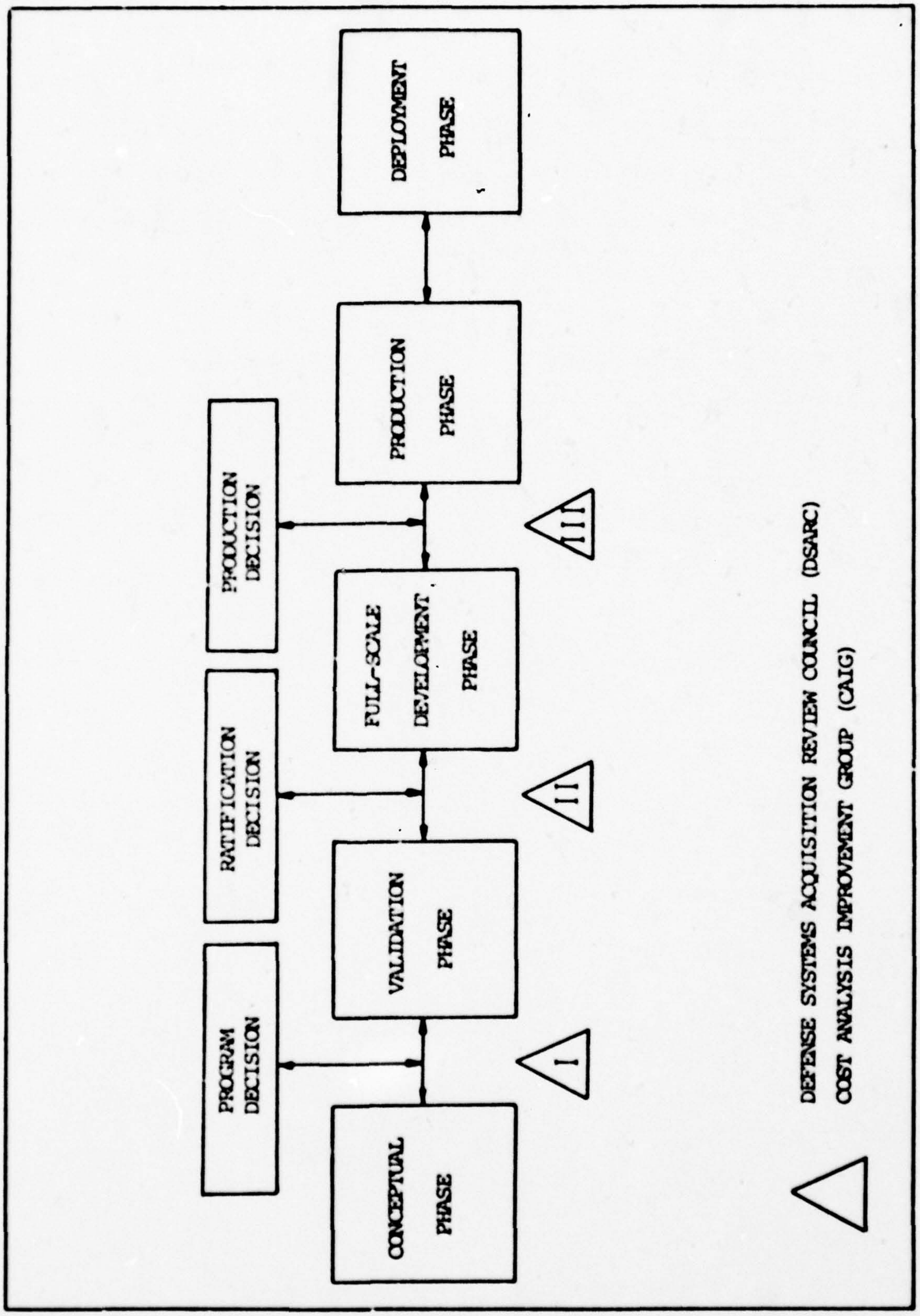


Figure 1. Weapons System Acquisition Process

selected provide cost information for budgeting and controlling the program (Ref 11:44). There are many techniques used for estimating system or program costs. Some of these techniques are discussed in the following section.

Methods Used to Estimate Costs

The range of cost estimating techniques consists of intuition at the one extreme to a detailed look at labor and material cost standards obtained from micro level time studies of worker actions. The three techniques that will be duscussed in the following paragraphs are the analogy, industrial engineering, and statistical approach (Ref 21:1).

The analogy approach consists of the estimator analyzing similarities and differences between some previously existing system and the one under study. The estimator then applies judgment to obtain the final cost estimate. Analogy is an approach used by government cost analyst to provide a rough check of an estimate made using another technique. Analogy has been used in private industry as the primary technique for obtaining estimates of costs. Estimators frequently use analogy when a firm moves into a new product line where no previous cost experience is available (Ref 21:1). An illustration of this situation occurred in the 1950s when many of the aircraft companies were trying to obtain contracts to build ballistic missiles.

These companies used analogies between aircraft and missiles to provide estimates of cost for a missile program. Douglas Aircraft used the method to obtain estimates for the Thor missile from its experience with the DC-4 transport aircraft. Good results were obtained. Later, when analogy was used again to obtain estimates for new missile systems using the Thor experience, the results were not as successful. The major drawback of this approach is that analogy relies heavily on judgment and experience (Ref 21:7).

Industrial engineering estimates are obtained through the detailed examination of work at the micro level and aggregating the many separate detailed estimates obtained into a total. The first steps in performing the micro level study requires the estimator to look over the set of engineering drawings and determine the engineering, tooling, and production operations required to produce the item. His analysis will specify the amount of labor and material required to be used. Where standards have been established the job of the estimator is simplified. He simply multiplies his estimates of labor and material by the applicable standards to obtain cost estimates for each component. These cost estimates are then aggregated into a total cost estimate. In the event that standards have not been established, a detailed study of each worker operation must be performed to specify the most efficient method for performing that task. Each task, using the most efficient method of performance is timed. The times required to accomplish each task are then aggregated into a total for that operation to determine a standard time. Because this technique is an in depth analysis, it is generally not used in the aerospace industry (Ref 21:2). The procedure requires more personnel, time, and data, then other methods. Also, the industrial engineering technique may be less accurate than estimates acquired from statistical methods. A reason for reduced accuracy is that small errors introduced in detail estimates can compound and result in a large total error. For example, if cost estimates of the elements, either manufacturing labor or material, are in error and overhead is a percentage of the aggregate of both, the total estimate will be in error by the amount of the error in the element plus a percentage of that error. Another reason that estimates acquired through statistical methods are better than estimates obtained from an industrial engineering approach may be due to the significant variability of factors assumed to be standard using an industrial engineering approach. Variability is frequently experienced in the aircraft industry and is caused by the factors of changing design, limited production runs of like models, and frequent unexpected production rates. These factors make it very difficult for standards, used in the industrial engineering approach, to be established or relevant throughout a program (Ref 21:6).

> The effect of these factors can be represented statistically by the learning or progress curve so characteristic of this industry. One set of fabrication and assembly modes is succeeded by more efficient production functions, which lower the total labor requirements (Ref 21:6).

The Department of Defense has found that statistical methods provide the ability for independent and objective cost estimates. They are not forced to rely completely on estimates provided by the industries, which tend to be optimistic and reflect the motivations of the preparers (Ref 12:55). The statistical approach discussed in this section is concerned with estimates obtained through use of a CER. Statistical estimating is sometimes defined as statistical extrapolation from cost or commitments experienced on the job to provide an estimate of the costs of the next unit or lot, or costs at the completion of the job (Ref 21:2). The latter defined technique will be discussed later as a method used to provide estimates from early actual cost data.

The statistical approach can be used to provide the required cost estimates necessary for long range planning through contract negotiation. Since the technique can be used for a variety of situations, it may vary in form for alternative applications. The variation in the application depends on the purpose for which the estimate will be used and the available data. For example, for long range planning the CER used may provide highly aggregated cost estimates since detailed knowledge may not be available or desired. In this phase of a program all that is required is an estimate of total cost. As the program moves closer to contract negotiation more detailed CERs would be used. These CERs would provide estimates of costs by categories. A set of categories that many organizations use for collecting cost data for accounting purposes is by functional category, e.g., engineering, tooling, labor, material, and overhead. This cost information provides management with budget and control targets to ensure efficient program performance (Ref 21:89).

When the statistical technique of regression analysis is applied to historical airframe cost data, which reflects learning-curve assumptions, CERs can be developed which provide "good" unbiased and reliable estimates. The subject of regression analysis and learning-curve theory will be discussed in more detail in Chapter II.

Statement of the Problem

The cost of complete systems with support equipment has jumped by a factor of seven during the period 1945 to 1955. From 1958 to the present the cost of complete systems continue to increase. Rising costs contribute to the current operating and support systems expenditures and absorb a greater portion of the total defense budget. Since the budget has been relatively fixed in recent fiscal years, less funds are available for concept and engineering development of new systems, let alone for development and production (Ref 4:26,28).

Due to rising costs and fixed budgets, greater responsibility has been placed on the Secretary of Defense, DSARC, DOD, and program managers for their decisions concerning the efficient expenditure of budget dollars to provide effective benefits (an acceptable defense posture). When operating under fixed budgets, the managers of the funds must ensure the best techniques of cost estimation and available information are used to forecast and control spending. If this is not accomplished, overruns will continue to be a DOD problem. As overruns absorb a greater portion of the fixed defense budget, funds available for development of other systems or procurement of additional systems will be limited (Ref 4:26,28).

Due to increased costs and competition for limited resources, the acquisition of airframes represents a significant expense to the Government of the United States. Early in the planning stages of acquisition, parametric methods are used. As development units or prototype units are completed, actual cost data becomes available for use in projecting costs of full-scale production. Projected costs for future production lots are also needed by the program manager for budgeting and controlling future expenditures. To accomplish budgeting and controlling, current methods require accurate determination of the components of development and early production lot costs that are recurring and the learning-curve slope. The current methods result in decreased accuracy due to the errors created by cost reporting systems to identify recurring and nonrecurring costs consistently and biases in selection of a learning-curve slope. A reliable statistical technique needs to be developed to estimate costs of succeeding lots of airframes for a new program using the available early actual cost data for the new program and historical airframe cost data.

Objectives

The primary objective of this thesis is to develop a model that provides reliable estimates of the next lot of airframes for a program using the prior actual cost data that is available. The model will hopefully provide "better" estimates than now can be obtained by application of current methods used to project future lot costs from observations on development units. The second objective is to obtain a confidence interval for the estimate obtained from the model. Another objective will be to compare the results obtained from the Bayesian equations with estimates obtained using several current methodologies. The comparisons should provide an indication of the value of the Bayesian methodology as an estimating tool.

Methodology

The RANDOM CERs developed by Marcotte in his thesis (Ref 19) were selected for the application of a Bayesian approach for updating CERs because the RANDOM CERs proved to be statistically "better" estimators than CERs developed using other currently published methodologies. The RANDOM CER was developed from a mixed linear model which considers two error terms.

$$Y = XB + u_j + \epsilon_{ij} \qquad (1)$$

where

Y = dependent variable of airframe cost.

X = independent variables of weight, speed, and quantity.

u = error due to a particular airframe.

ϵ = error term across all lots of airframes.

To estimate the parameters of the RANDOM CER the concept of Henderson's Method 3 for fitting constants was used along with the concept of Generalized Least-Squares (AITKEN) Estimators. Generalized Least-Squares (GLS) is a methodology that provides estimates of the parameters of a general linear model having correlated error terms.

The primary thrust of this research effort is to apply a Bayesian philosophy to the problem of estimating future lot costs of an airframe program. Bayesian philosophy allows usage of all information available, both the current observations for the new airframe program and the past historical data on all other programs. A methodology discovered in the literature that is applicable to this problem is a Bayesian approach to multiple regression presented by Sasaki (Ref 30). The methodology allows regression coefficients from a previous experiment to be updated as new information becomes available on a new experiment. The updated coefficients are the posterior estimates of the regression coefficients. The same value for the posterior coefficients obtained using the Bayesian methodology can be obtained by adding the data from the previous experiment to the data provided by the new experiment and performing regression analysis on the combined data set.

The important result of the Bayesian methodology presented by Sasaki is the equation developed to estimate the posterior variance, $\hat{\sigma}^2$.

$$\hat{\sigma}^2 = \frac{\sigma_1^{\,2}\gamma_1 + \sigma_2^{\,2}\gamma_2}{\gamma} \tag{2}$$

where

$\sigma_1^{\,2}$ = variance of the first experiment, prior information

$\sigma_2^{\,2}$ = variance of the second experiment, current information

γ_1 = degrees of freedon of the first experiment

γ_2 = degrees of freedom of the second experiment

γ = total degrees of freedom $(\gamma_1 + \gamma_2)$

Equation (2) weighs the variance of each experiment by the number of degrees of freedom for each experiment and divides the sum by the total number of degrees of freedom (Ref 30: 472).

Since the RANDOM CERs were developed using historical fighter airframe cost information, the estimated parameters are considered prior information for the application of a Bayesian approach to updating airframe CERs. The lot cost data observed for a new airframe program are considered to be current information from a new experiment. Since regression analysis on the combined data sets provide the same regression coefficient values as the posterior regression coefficients resulting from the Bayesian methodology presented by Sasaki (Ref 30), regression analysis will be performed using combined data sets to obtain the posterior coefficient estimates $(\hat{B})$.

Due to the assumption that correlation exists among the observations, GLS techniques will be employed to estimate the coefficient parameters. The posterior variance will be estimated using equation (2). The equation requires estimates of $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ which will be obtained from independent regression analyses on the historical cost data and the new program data.

Because correlation exists among the observations of a new program, prediction requires a special treatment using the GLS methodology presented in Johnston (Ref 14: 208-213). The posterior CER developed requires the addition of an adjustment factor which takes into account the correlation between the observed lots and the lot to be predicted. Details of the GLS methodology are presented later in the thesis as outlined in the following section.

Organization

The introduction to this research has provided the reader with the statement of the problem, objectives of the research, and a brief discussion of the methodology employed. Chapter II discusses the techniques and theory used to develop CERs for estimating airframe cost. Summaries of several recent studies conducted which develop airframe CERs and two studies which attempt to provide a technique for projecting future costs from early actual costs and provided as background of the state-of-the-art. Chapter III is a discussion of the data bases used to conduct this research, and a discussion of the methodology used by the writer to develop CERs which predict unit average cost in terms of weight and speed. Chapter IV addresses the methodology employed and application for developing airframe CERs. Chapter V presents the Bayesian CERs developed and comparisions of the Bayesian CER predictions with estimates obtained from several other CERs.

The comparisons are made to determine if the Bayesian methodology provides a "better" estimate of future lot cost than other methodologies. The last chapter, Chapter VI, summarizes the research effort, discusses the conclusions, and presents recommendations.

II. Background

This chapter presents background information that relates to development of CERs and updating estimates. The chapter is broken into two sections.

The first section of this chapter is intended to provide the reader with an understanding of the two basic concepts used to obtain airframe CERs. The first concept is the phenomenon commonly referred to as the learning-curve. The second concept is a statistical method that estimates parameters which provide a "best fit" straight line through the observed data. This concept is referred to as regression analysis. The last concept presented is the mixed linear regression model containing fixed and random effects. The RANDOM CERs used in this study were developed using a mixed linear model. The latter part of this section summarizes several studies which use various methodologies, employing the techniques discussed above, to develop airframe CERs. Predictions made using the Bayesian CERs will be compared with predictions made using the CERs developed from these various methodologies to determine the predictive capability of the Bayesian approach.

The second section of the chapter presents the results of a literature search for statistical approaches which update cost estimates. Two studies were found. One study develops a formula to update current estimates for prediction of future procurement cost. Some of the parameters require the use of subjective information. The other study attempts to argue for a modified learning-curve developed using discrete uniform probabilities and a cubic function.

Development of CERs

This section includes a discussion of learning-curve theory and regression analysis as these concepts apply to development of CERs for predicting airframe costs. An iterative method developed by Henderson, modified by Thompson, and used by Marcotte in his thesis (Ref 19) to obtain airframe CERs, is also discussed. The CERs to predict airframe costs, developed by Levenson, et al in their report, Cost-Estimating Relationships for Aircraft Airframes (Ref 18); Marcotte in his thesis, Airfraft Airframe Cost Estimation Utilizing a Components of Variance Model (Ref 19); and by the writer, using the methodology in the Large study (Ref 17), are used to provide estimates for comparison to determine predictive capability of the Bayesian approach. Each of the studies are summarized for the reader in this section.

Learning-Curve Theory. The primary reason for discussing learning-curve theory is to provide the reader with understanding of a method used commonly today by the aircraft industry and by government cost estimators to project cost estimates of follow-on production from early actual cost. The phenomenon of decreasing costs with increasing quantity of production was observed by

T. P. Wright in 1936. This phenomenon is known as the learning-curve. Wright identified the following elements that affect the cost of aircraft.

1. Design factors including type of material used.

2. Consideration of tooling and production methods as the airframe is designed.

3. Extent of engineering changes once into established production.

4. Size and weight of the airframe.

5. Quantity of airframes built.

His primary interest was on the effect of increasing production on the cost of labor, material and overhead. Wright plotted labor cost data versus quantity of airframes. From the shape of the curve drawn, it became evident that the curve was of the form generated by the formula $C = X^b$ (Ref 37:122-125). This curve was called a learning-curve because analyst felt the learning process was responsible for the phenomenon of decreased cost either in man-hours or dollars as quantity produced increased. Since then, analyst have determined that many factors are enumerated in the Military Equipment Analysis handbook published by RAND, and are listed below.

1. Job familiarization of workmen, which results from the repetition of manufacturing operations.

2. General improvement in tool coordination, shop organization and engineering liaison.

3. Development of more efficiently produced subassemblies.

4. Development of more efficient parts-supply systems.

5. Development of more efficient tools.

6. Substitution of cast or forged components for machined components.

7. Improvement in overall management (Ref 21:94).

With the revelation that these and other factors are responsible for the phenomenon observed, many new titles have evolved into common use, "experience curve," "time reduction curve," "progress curve," and "percent improvement curve" (Ref 20:I-1).

The learning-curve, $C = X^b$, illustrated at the top of figure 2, is graphed on arithmetic coordinates. When the curve is plotted on log-log coordinates, illustrated at the bottom of figure 2, the relationship of cost to quantity produced changes to a linear relationship. Basically, each time the quantity of goods produced doubles the cost per item decreases by a constant percentage of the previous cost of the item (Ref 21:93). For example, an item cost $100 per unit when 20 units are produced. As production increases to 40 units the cost decreases by 20 percent, to $80 per unit. If cost again decreases 20 percent when 80 units are produced, the curve would be an 80 percent learning-curve.

The formula that is generally used to express this particular relationship is

$$Y = AX^B \tag{3}$$

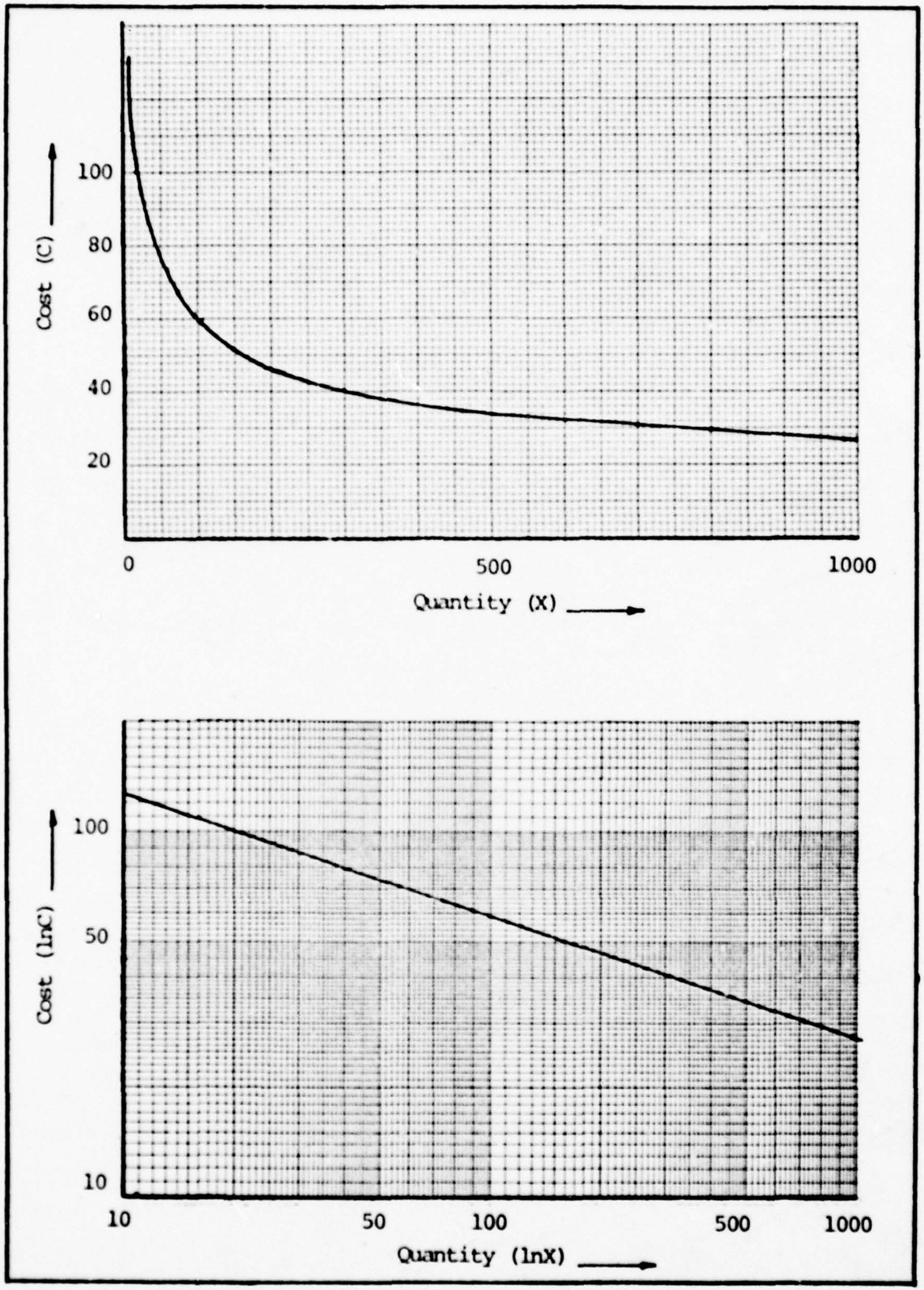


Figure 2. Eighty Percent Learning-Curve (Ref 21:95)

where,

X = the number of the unit produced.

Y = the number of direct man-hours or dollars required to produce the X^{th} unit.

A = the number of man-hours or dollars to produce the first unit.

B = log s/log 2 where s equals the learning factor (Ref 8:488).

This function is known as an exponential (log-linear) equation. The dependent variable (Y) gets smaller as the independent variable (X) increases. Both A and B are constants where A can assume any positive value and B is constrained to a negative constant between 0 and -1 (Ref 25: 12).

The value of A determines the value or magnitude of vertical height of the dependent variable (Y). The constant exponent B controls the rate of decrease of Y as X increases. As B approaches the value of zero, Y approaches a horizontal line, A units high and tends to decrease very little for increasing values of X. Conversely, as B approaches -1, the rate of decrease of Y grows larger (Ref 25:13).

The first acceptance of learning-curve theory was in the aircraft industry during World War II. Management desired a method for predicting cost, manpower requirements, and setting prices in an industry facing many changes and increasing complexity (Ref 20:I-4). Analyst discovered that the following learning-curve assumptions were applicable to this industry:

1. The time required to perform a given task decreases each time the task is performed.

2. The unit time decreases at a decreasing rate.

3. The time reduction will follow a specific and predictable function, such as an exponential (Ref 9:100).

The learning-curve method continues to be used in the aircraft industry to provide estimates for planning and control. Some of these usages, as pointed out by Asher, include cost estimating required for contract negotiations, budget estimations, and preparation of airframe production schedules (Ref 5:4). The most frequent use of learning-curve theory is to provide cost estimates. The following are samples of the various situations where learning-curve theory can be used for estimating costs:

1. New aircraft proposals. Cost estimates are required to determine pricing during contract negotiations. Past production experience on a similar item provides insight into the type function that represents the learning phenomenon and the slope of the curve.

2. Follow-on production. Once an airframe has been in production, learning theory is applied to the available actual production cost data in order to estimate future product prices. This may be required for price negotiations or purchase of additional lots.

3. Changes. When modifications are proposed, the analysts use basically the same procedures used in determining new airframe costs, but have the advantage of current airframe production data. The cost estimate for the change is evaluated in terms of the additional construction required to be made to the present airframe.

4. Unsolicited proposals. Along with competitive bids based on learning-curve theory, the company commits funds for development aircraft based on estimates provided by learning-curve theory. These development aircraft or prototypes which are produced to provide a basis for unsolicited proposals. The resulting proposal will hopefully be evaluated and approved by the government, and lead to follow-on contracts (Ref 6:16,17).

When production has not started, the estimation procedure using learning-curve theory becomes enlightened guess work. Several options are open to the analyst. One option has been used for previous applications in the industry. Another option is for the analyst to assume a learning-curve percentage that is applicable for the same or similar products. The last option is to analyze the similarities and differences between the proposed startup and the previous startups to develop a new learning-curve percentage which appears to best fit the situation (Ref 8:490).

Two common methods are used for applying learning-curve theory to cost estimating. The first method requires an estimate of the cost of producing the first unit. An accurate first unit estimate is difficult due to the many sources of variation in producing the first unit. An assumed learning-curve percentage is then applied to estimate the costs for following units. The second method establishes cost at some unit well along in the production curve process (i.e. 100th, 200th, etc.); then an assumed learning-curve percentage is applied to determine the cost of the preceding and following units. The "number one cost" using the second method is defined as the theoretical cost of producing the first unit. This first unit cost is termed theoretical since actual first unit costs are rarely equivalent with the costs estimated by the curve. Errors are generally compounded by both methods due to the selection of an assumed learning-curve percentage (Ref 20:VII-1, VII-2).

Many factors influence the assumed learning-curve percentage selected by an analyst. One factor is the amount of initial investment in planning and engineering. The more money spent initially in planning and engineering reduces the number of changes required to be made later in the production process, design, and labor force; therefore, the rate of improvement is generally less than the converse situation. Low rates of improvement are associated with high percentage learning-curves. Another factor which also effects the selection of a learning-curve percentage is how the work will be accomplished, either by man or machine. Labor intensive production systems have high rates of improvement, while machine intensive systems have low rates of improvement. A fact related to this is that different types of labor generate different percentages of improvement. Assembly line labor shows a very high rate of improvement. The fabrication worker has a lower rate of improvement. Many times the worker is not responsible for the particular rate of learning, but the nature of the work. For example, the speed of a job depends more on the equipment than on the skill of the operator. The rate of learning improvement is reflected only in lessening the time required for set-up and maintenance. If these two items cannot be reduced significantly, the skilled operator of the equipment has a very low rate of learning (Ref 20:I-8,I-9).

The reliability of the cost estimates from the application of learning-curve theory depend on the estimate of first unit cost and the assumed learning-curve percentage. Where production has begun, one source of error in the application of learning-curve theory to early actual cost is eliminated because the value of the "number one cost" is no longer an estimate but an actual value. Other assumptions which are also important follow:

1. The same production condition exists for follow-on production as existed in the past.

2. Large changes do not occur in the labor force.

3. Major changes will not be made to the airframe (Ref 6:10).

If an error does occur in an estimate, many times the error can be attributed to one of the three assumptions not holding (Ref 21 and 6:118 and 11).

Regression Analysis. Regression is a statistical method that identifies patterns in historical or observed data. Regression provides a method to measure the relationship between two or more variables. There are basically two main objectives for performing a regression analysis. The first and most important to scientific research is for testing hypotheses. Because many hypotheses in the social sciences cannot be proved using deductive methods, the statistical method of regression is an invaluable tool to the social scientist for verification of hypothesis. The second objective is to make predictions of the values of variables (Ref 30:404). Development of cost estimating relationships to predict airframe costs as a function of one or more independent variables using past historical data is an example of this second reason.

Though it is desirable to predict a value of one variable exactly in terms of others, this rarely occurs. Regression analysis predicts an average or expected value.

The theoretical model underlying regression analysis is

$$E(Y_i | X_i) = A + BX_i \text{ for all } i \qquad (4)$$

where X_i are constants and Y_i are random variables with a conditional probability distribution. The assumption of a conditional distribution which is a set of random variables normally distributed, underlies regression analysis. The rational for this assumption is that there exist two types of errors in regression analysis. The first is the error in the variable which involves the measurement of Y. The other type is the failure to include certain variables that influence Y. The latter error is referred to as the error in the equation. The common assumption made is that there are no errors involved with the measurement of the variable, but only in the equation.

Each conditional probability distribution has a mean or expected value of Y_i which is expressed by

$$E(Y_i|X_i) = \mu_{Y_i|X_i} \quad (5)$$

and similarly the variance by

$$VAR\ (Y_i|X_i) = \sigma^2 \quad (6)$$

Regression theory further assumes the conditional distributions have equal size standard deviations, $\sigma_i = \sigma_j = \sigma$. In regression analysis, the regression curve is assumed to pass through the means, $E(Y_i|X_i)$, of the conditional distributions and to be linear. Since the individual value of Y is not exactly equal to the mean, the deviation is expressed by ε_i and must be added to equation (4) (Ref 30:405-407). The A parameter, a constant, is the Y intercept of the curve and the B parameter determines the slope of the linear relationship. For a curvilinear relationship B is referred to as the regression coefficient.

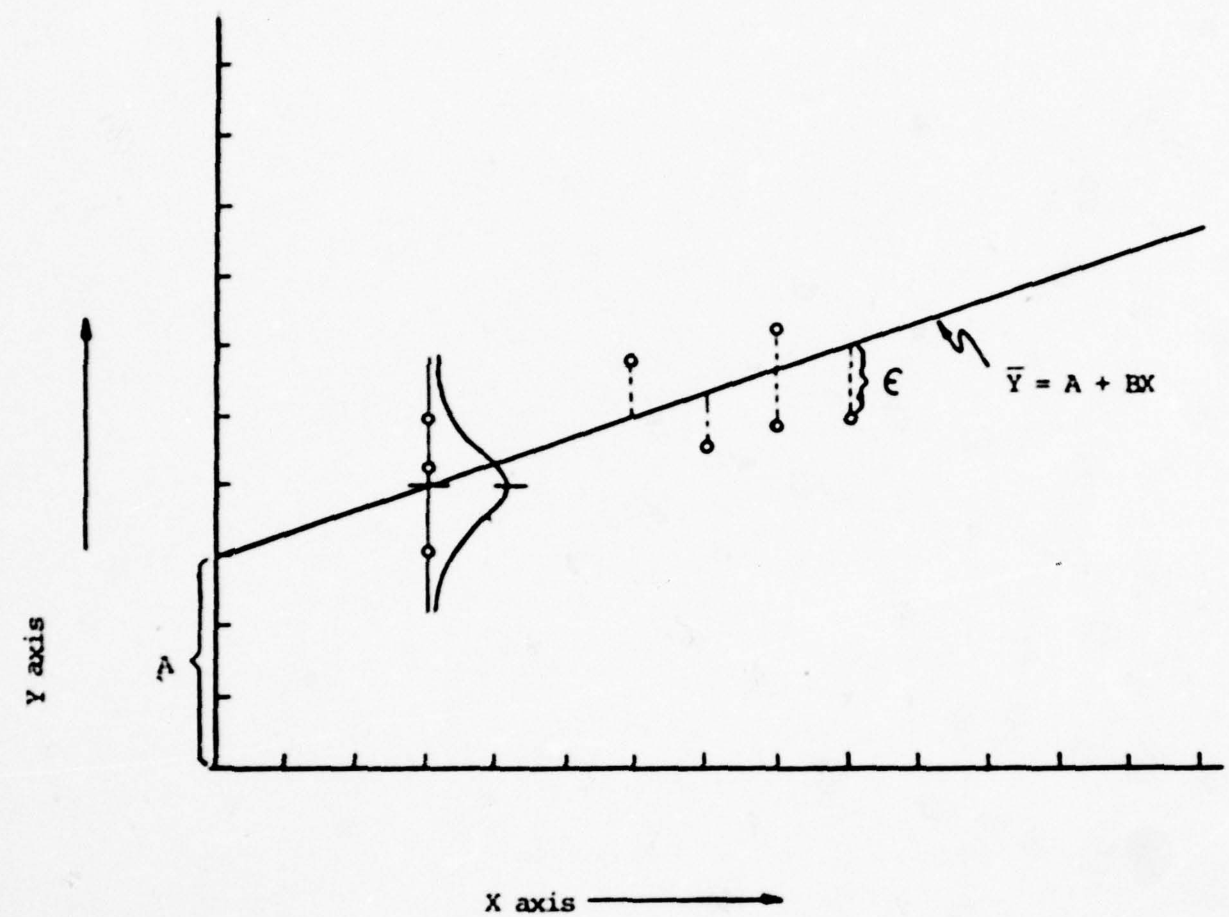
There are several methods for performing regression on observed data to obtain estimates for A and B. The most common method is least-squares, which minimizes the sum of the vertical deviations squared from the regression line, figure 3. The derivation of this technique will not be discussed here. If the readar desires information on this subject see Freund (Ref 11:358) or Sasaki (Ref 30: 404).

If logarithms are taken of both sides of the exponential learning-curve formula, equation (3), the formula is transformed into the form of a linear regression equation,

$$\ln Y = \alpha \ln e + \beta \ln X \qquad (7)$$

where, $\alpha \ln e = A$ and $\beta = B$

By performing multivariate regression analysis using explanatory variables such as weight and speed, more complex functions are developed which are used to predict the cost of prototype airframes, even when observed data is not available. Because CERs developed using multivariate regression contain explanatory variables which reflect the



o - Observation

ε - Difference Between Observed Value and Predicted Value

Figure 3. Illustration of the Relationship of the Regression Line and Observed Data

causal effects of cost, the multivariate CER has improved the accuracy of cost predictions.

Two widely accepted functional forms for predicting airframe cost are the (Ref 38:6),

Logarithmic Model

$$\text{Cost} = e^{a}W^{b}S^{c}Q^{d}\varepsilon \quad (8)$$

Exponential Model

$$\text{Cost} = e^{a}W^{b}S^{c}Q^{d}+\varepsilon \quad (9)$$

where,

a,b,c,d = regression coefficients derived from actual data.

W = Weight of the Airframe

S = Speed of the Aircraft

Q = Quantity produced

ε = Residual error term

The basic linear regression form of the equation is, $\ln \text{Cost} = a \ln e + b \ln W + c \ln S + d \ln Q + \ln \varepsilon$.

Fitting Constants Method. Fitting fixed effect linear models is often referred to as the technique of fitting constants because the effects are sometimes called constants. The general form of a linear model is

$$Y = X\beta + \varepsilon \quad (10)$$

Y = dependent variable

X = independent variable

ε = error term

Henderson (1953) developed a method, Henderson's method 3, for estimating the variance components which is based on the fitting of these models. This method is now commonly referred to as the fitting constants method. It uses reductions, R( )-terms, of sums of squares due to fitting the model and submodels. These reductions are then used to estimate the variance components by equating each computed reduction to its expected value. This basically is the same manner in which the sum of square terms are used in the least-squares technique of regression analysis. The β vector can be fixed, random or mixed (Ref 32:443,444). The fitting constants method provides estimates for the β terms as does least-squares, therefore this technique can be used to develop CERs.

Cunningham and Henderson (1968) used the fitting constants method on the following linear mixed model,

$$Y = X\beta + Zu + \varepsilon \qquad (11)$$

where,

Y = dependent variable

X = represents fixed effects

u = represents random effects

ε = error term

Equations to estimate the variances of the error term (ε) and the random effects (Zu) were developed using an iterative procedure based on maximum likelihood equations implicit in Henderson's Method. Thompson (1969) discovered an error in the presentation made by Cunningham and Henderson. The error resulted in the estimates of the variance components to be biased. Thompson then developed equations to provide unbiased estimates for the variance components (Ref 32:465). A detailed explanation of this method is presented in a book written by Searle, Linear Models (Ref 32).

Marcotte Study. This study investigated the application of a mixed fixed effects/random effects model, Henderson's Method 3, to historical airframe cost data of nine fighter type aircraft consisting of 33 lots. It was conducted as a result of an observation made by Womer, that lot costs of the same type aircraft are correlated with one another as compared to lot costs among different or new airframes. Figure 4, is a reproduction of the graph made by Womer of residuals versus weight which illustrates this point. Womer further argues that the total error between actual costs and the cost estimates obtained from the regression equation is made up of two components. One component of error is error due to a particular program, type aircraft. The other component of error is the error due to the regression across all lots and types of aircraft in the data base. This second component is the residual error normally obtained from a regression analysis (Ref 36: 10-14).

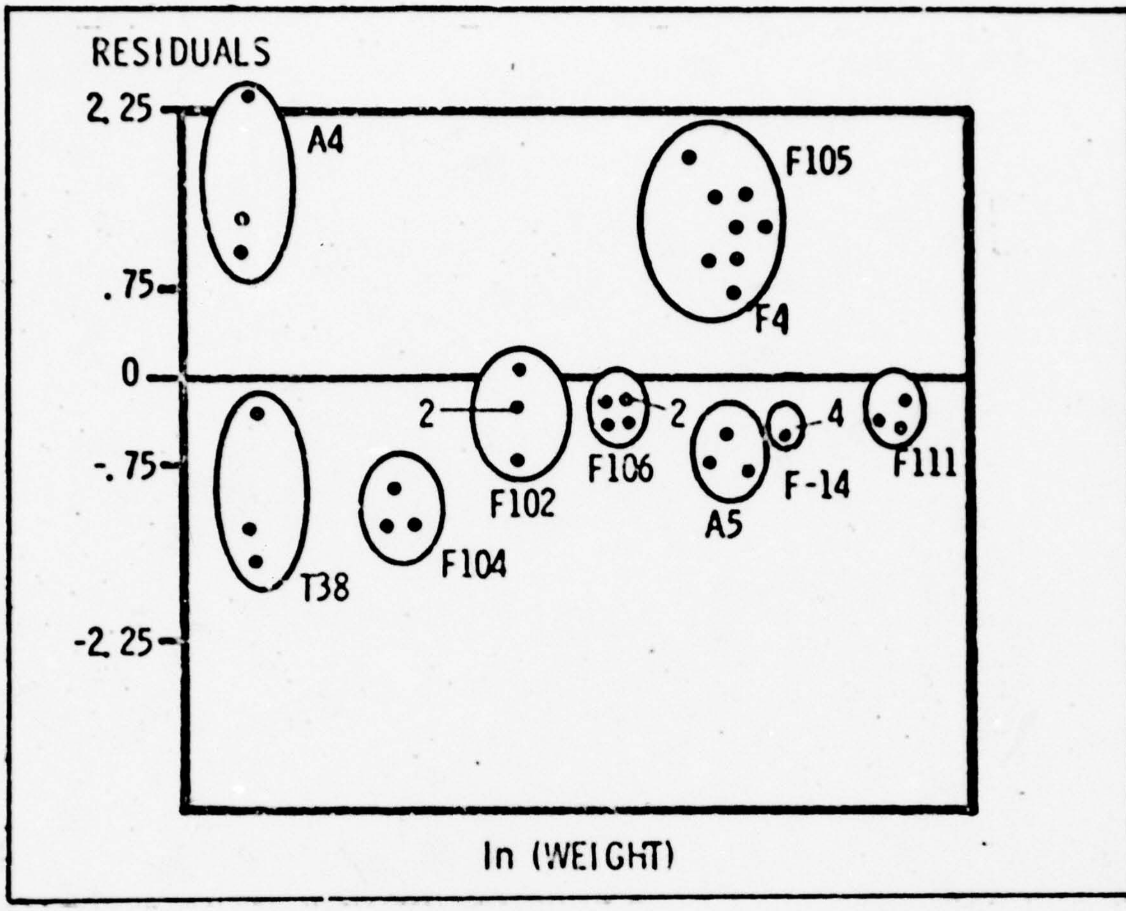


FIGURE 4* - Standardized Residuals Versus Weight (Ref 36:11)

* Multiple observations at the same location are indicated by numbers.

Marcotte (Ref 19) adapted the mixed model composed of random effects and fixed effects to airframe cost data and developed CERs to predict both airframe cumulative and marginal cost of total engineering, total tooling, recurring labor, and recurring material. In his presentation, the random effects (u) of the mixed model is the random error associated with a particular aircraft type. The residual error term (ε_{ij}) is the random error across all aircraft types (i) and all lots (j). The assumptions made in the study are that the errors are independent and normally distributed with zero means and variances (σ_u^2 and σ_ε^2) and the total error has a constant variance. The fixed effects (X) of cost is due to the variables of aircraft weight, speed, and quantity produced. The (X) matrix, 33 by 4, was constructed as follows:

$$X = \begin{bmatrix} 1 & \ln S & \ln W & \ln Q \\ 1 & . & . & . \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \end{bmatrix}$$

The random effects, Z matrix, was constructed using columns of zeros and ones. The one represents a lot of a particular aircraft type. Since nine fighter airframe types were used in the data base, nine columns were constructed as follows:

$$Z = \begin{bmatrix} 1 & 0 & 0 & . & . \\ 1 & 0 & 0 & . & . \\ . & 1 & 0 & . & . \\ . & 0 & 1 & . & . \\ . & . & 1 & . & . \\ . & . & 0 & . & . \end{bmatrix}$$

When Z was multipled to the column vector of u, the result was the random error due to a particular airframe program.

The CERs developed by this technique, referred to as the RANDOM technique, were compared with CERs developed by Marcotte using the same data base of nine fighter aircraft and the methodologies employed by Large, et al (Ref 17), and Levenson, et al (Ref 18). The significant conclusions drawn by Marcotte from the comparisons were that the RANDOM technique provides a "better" estimate of cost for the first and second lot than did the other CERs developed. The statistics, squared correlation coefficient (R^2), estimate of the variance of the residuals (${\sigma_{\varepsilon}}^2{}_1$), and t-ratios, of the CERs developed from the RANDOM techniques were consistently better than the statistics of the CERs developed using the other methodologies (Ref 19:64-68).

Levenson, Boren, and et al Report. A RAND study performed in 1971 reported the results of statistical research accomplished involving development of CERs for airframes. The CERs developed were to be used to estimate both development and production costs of aircraft airframes. They were mainly to be used for long range planning. Separate equations were developed which provide estimates for engineering, development support, flight test operations, tooling, manufacturing labor, manufacturing material, and quality control. Additional equations were developed for prototype estimation.

The data base used for the study consisted of 29, primarily aluminum, post-World War II aircraft. The physical and performance characteristics range in weights from 5,000 to 113,000 pounds and maximum speeds from Mach 0.5 to Mach 2.2. The cost data was separated into appropriate categories, and adjusted to ensure comparability and compatibility.

The CERs were derived through application of multiple regression techniques to the historical data. The equations were expressed as exponential forms. The three explanatory variables that best explained the variations in cost were aircraft quantity, maximum speed, and weight. However, an addition of other explanatory variables to these were required for some categories of cost. The tooling equation used production rate. To explain the cost of flight test operations, this equation required the addition of a variable for the number of flight test vehicles needed (Ref 18:v).

Large Report. This report was performed by RAND in 1975. The research conducted developed generalized equations for predicting development and production costs of airframes using two variables, aircraft weight and speed. This information is usually defined and available early in a program. Separate equations were developed for the following cost elements:

1. Engineering
2. Tooling
3. Nonrecurring manufacturing labor
4. Recurring manufacturing labor
5. Nonrecurring manufacturing material
6. Recurring manufacturing material
7. Flight test operations
8. Quality control

Equations for estimating total program cost and prototype development cost were also developed.

The data base used in this research consisted of cost data on 25 military aircraft from 1953 to 1970. This study attempted to eliminate or reduce the variances caused by differences in construction and manufacturing techniques used prior to 1953. (Many of the earlier studies derived CERs using aircraft developed as far back as 1946.) The sample of aircraft ranged in weight from 5,000 to 279,000 pounds and in maximum speed from 300 to better than 1300 knots (Ref 17:v).

The technique of multiple regression was the foundation of the methodology employed. In order to obtain input values, the cumulative total hours or dollars were plotted against cumulative quantity. Lines were drawn between plot points. From the line drawn, cost values were picked off for quantities 25, 50, 100, and 200.

The reason for examing cost at several quantities was to determine if the cost-quantity curve was segmented, rather than linear. The conclusion was that the curve is linear beyond the 24th aircraft and nothing was to be gained by using a segmented curve. The CERs developed can be used to predict the cost for either the 25, 50, 100, or 200 cumulative units.

A stepwise least-squares procedure was used initially. This method determined the explanatory variables that are statistically significant in explaining cost. Due to the low predictive value, many of the explanatory variables were eliminated leaving four or five variables remaining to be examined. After checking for the logic of each of these variables, many were discarded. It was concluded in this study that even with their deficiencies, weight and speed are dependable predictors of airframe costs. Also, the logarithmic model CER showed a better distribution of the residuals than the exponential form (Ref 17:16-17).

> This is consistent with current belief at RAND that (1) the error distributions for cost data tend to be more constant over the range of data in the logarithms than in the actual (raw) values, and (2) the criterion of percentage (relative) errors is more appropriate than on of actual errors. (The logarithmic regression minimizes relative errors rather than actual errors as in the power regression.) (Ref 17: 17).

Projecting Cost from Early Actuals

From the literature search performed, only two statistical studies were found that attempted to improve estimates.

A RAND report by Robert Summers, Cost Estimates as Predictors of Actual Weapon Costs (Ref 33), developed statistically a debiasing equation that would be useful for adjusting current estimates in predicting future procurement cost. The other study by Van Puryear, New Progress Functions - With Probability Modifiers (Ref 26), analyzed the log-linear learning-curves as they exist, and then modified them using discrete uniform probabilities and a cubic function.

In a discussion of cost estimating techniques with Chuck Samson of the Aeronautical Systems Division, Comptroller, Cost Analysis Directorate, Cost Estimating Division (ACCC), several methods currently used for projecting estimates from early actual cost data were discussed. These techniques along with summaries of the two studies will be presented in this section.

Summers Report. This report is a RAND report performed in 1962. The report consists of a statistical study of a sample of military cost estimates. An objective was to decrease the uncertainties about the interpretation and use of these estimates. The sample of cost estimates consisted of estimates made for major hardware articles in 22 weapons systems. These estimates were compared with the actual costs to assess their accuracy as predictors. According to Summers, the grosser differences disappear when the estimates are adjusted for the actual quantities of the end item procured and take into account the changes in price levels due to inflation (Ref 33:v).

The research performed found that even when the adjustments were made, great variablity still existed, ranging from 15% to 150% of actual cost. The study identified situations where the variability is likely to be large. Summers developed a "debiasing" equation so the estimates, although still variable, would be more likely to be low than high. The debiasing formula developed using multiple regression techniques illustrates how low a cost estimate is likely to be. The formula follows (Ref 33:8):

$$F = 11.929 \left\{ \exp \left[.097t - .032tA - .311A + .015A^2 + .008L - .075(T-1940) \right] \right\} v \qquad (12)$$

F = ratio of actual cost to adjusted estimate.

t = the timing of the estimate within the development program expressed as a fraction of program length.

A = the degree of technological advance required in the program (on a numerical scale).

L = the length of development period (in months).

T = the calendar year.

v = residual error of the regression.

When an original estimate was multipled by a debiasing factor (F) of two or three, the new estimate still has variability associated with it. Because the new estimate was no more likely to be low than high, the revised estimate provided a higher degree of confidence. Summers said, "It will almost certainly be closer to the true cost than the original estimate (Ref 33:10)."

The important parameters of the debiasing formula are t, A, and L. Greatest uncertainty exists in the value for A. This is due to the subjectivity in the establishment of the numerical scale and where a particular program is on the scale. The scale used was developed from a survey of experienced RAND engineers, which rated the magnitude of improvement in the state-of-the-art for each program in the sample.

Puryear Study. The purpose of this research was to develop an improved technique that would provide management with a tool in the Acquisition Process for determining reliable and accurate cost estimates. Puryear reviewed 2388 recourds of learning-curve experience on a variety of Army equipment. The outcome of the record review was that for 94% of the items purchased the learning-curve percentage was 90%-100%. This fact led to the logical conclusion that, "the greater the amount of projected improvement of a given improvement function the less the chance (probability) of achieving that reduction" (Ref 26:11). Puryear then establishes a relative probability curve. The curve consists of a linear graph from a probability of 1.0 where no reduction in cost occurs, 100% learning-curve percentage, to a .05 change of achieving a projected productivity or learning-curve percentage of 75% (Ref 26:11). "Equal Chance" in probability theory states, that for a large number of trails there is equal probability for achieving a particular trial. The "Equal Chance" theory was applied to 11 learning-curves, i.e. 100%, 98%, 97%, ... 90%, and each curve assigned a 1/11 chance of being achieved. The 1/11 probability was then weighed by the relative difficulty of achieving the projected learning-curve, supported by the sample data. This weighing modifies the latter linear portion of the function $y = ax^{-b}$ and shifts it to a higher position on arithmetic grid. The beginning part of the learning-curve is modified by a cubic function: $y = ax + bX^2 + cx^3 + d$ (Ref 26:14). Puryear speculated that the cubic function was applicable for the following reason:

> The expected production rate, for prorated investment, is a slow initial rate, a debugging and management improvement phase which has a sizable increase in productivity, followed by a reducing rate of productivity as the production continues (Ref 26:14).

Techniques Used by Cost Estimators to Project from Early Actuals. In a conversation with Chuck Samson, a cost estimator in the Cost Estimating Division of the Comptroller for Aerospace Systems Division, several techniques used to project cost estimates for follow-on production from early actuals were discussed. One method is regression analysis, provided there is enough early actual data available. This technique is considered "better" since it averages the variations statistically and is not subject to the implicit biases injected by use of the other methods which require judgment. Another way to obtain the learning-curve slope is to visually inspect the observed actual data and simply draw a straight line approximation. For rough estimates this technique is often used according to Bannon (Ref 6:6). Bannon gleaned the following from conversations with estimators employed by a well known airframe contractor.

> Actually, this method is not as 'rough' as one would expect. A good analyst has knowledge of not only past performance but of future production considerations and requirements. He is able to weigh the effect of changes in the labor force, the airframe, the production run, etc., and can use his expert opinion to influence his best fit of a curve to this data (Ref 6:6).

A third approach, provided the company contracted to perform the development work on an airframe has had a recent program which is similar to the current program, uses the learning-curve slope and program experience from this past program to provide the estimate for follow-on units. The last alternative technique available is to use an industry wide learning-curve. This latter technique has many fallacies as pointed out by Chase and Aquilano (Ref 8).

> In any case, while a number of industries have used learning-curves extensively, uncritical acceptance of the industry norm (such as the 80 percent figure for the airframe industry) is risky, and therefore, an analysis of the similarities and differences should be undertaken even though it may ultimately lead to the industry improvement percentage (Ref 8:490).

Some reasons for the disparities between industry rates of learning and the individual firms rate, are the differences in operating characteristics stemming from equipment, methods, product design, plant organization, and the procedural differences in the determination of the learning percentage. An example of the latter is an industry rate for a single product or a product line which hinges on the manner in which the data is aggregated (Ref 8:490).

McDonald also criticizes the use of an average industry curve. Each company and product is different which causes different rates of improvement for each firm and sometimes each product. Firms with a high percentage improvement curve of 95% have maintained with some justification that the production was properly planned from the very beginning and that a 70% curve indicates poor initial planning and execution (Ref 20:I-7).

Summary

The first section of this chapter discussed the theory used for developing airframe CERs: learning-curve theory, regression, and the fitting constants method. This discussion was included to provide the reader with the background theory of airframe CERs. The section also summarized reports of various state-of-the-art methodologies used to develop airframe CERs. The CERs developed using the methodologies discussed will be used for evaluating the predictive capabilities of the Bayesian approach.

The second section summarizes studies which investigated statistical approaches for updating cost estimates. Both studies were the result of an in-depth literature search. The studies are presented to provide the reader with background on the state-of-the-art approaches to updating cost estimates since this problem relates directly to the stated problem of this thesis. Also included in this section was a brief discussion of other techniques used by the cost estimator to forecast from early actual costs.

III. Data Base

Of the many hours spent in the development of a CER more time and effort is expended on the collection and adjustment of the data base than any other aspect (Ref 10:130). Any extensive empirical work, such as development of a CER, requires a collection of suitable data as the first step in the analysis. The data base is a source of two problems which surface in the development of a CER. According to Johnston, "the first major difficulty is the extreme paucity of published data on costs and outputs, a difficulty that is aggravated by the secrecy that often surrounds unpublished data on these variables (Ref 15:2)". The second major problem is to insure the data is relevant and conforms to the theories used in the development of the CER. The latter problem may require adjustments to the data (Ref 15:2).

This chapter begins with a discussion of data bases in general. The discussion concerns adjustments and the difficulties of separating costs into recurring and non-recurring categories. Separation of costs into these two categories is an essential requirement to accurately estimate follow-on lot or unit costs based on early actual costs. Following this is a section which discusses the data base used for development of the Bayesian CERs. The sources of the data and the various categories of costs which will be analyzed using a Bayesian approach are described. The last section of this chapter presents the methodology employed to obtain marginal unit CERs for the 50th and 100th unit. These CERs use only the explanatory variables of weight and speed. The same methodology was used in the Large study to obtain the cost of a particular unit airframe from information on cost accumlated by lots.

Data Bases in General

The development of a cost estimating procedure requires at least three types of historical data. The first type is resource data which includes expenditures for material and labor hours. Resource data may have either of two classifications. One classification is end-item, which means the cost for resources is collected by system, subsystem, component, and part. The other classiciation is functional category, i.e. engineering, tooling, labor, material, and overhead. The second type of data required is descriptive data. Descriptive data are the various performance or physical characteristics of the item that have a direct relationship to the cost of the item. For airframes, past analyses have determined that weight, speed, and quantity are "good" cost-explanatory variables. The third type of data required is program data. This data is information concerning past development and production history of a particular hardware program. Program data includes acceptance data, significant milestones, production rates, ECPs, and contract lot sizes. This data provides insight into the cause for the fluctuations and variations in the program as compared to other programs or industry historical data (Ref 21:13-16). For major programs, program data can be obtained from the Cost Information Report (CIR), a cost report required by the government from the contractors. Many times complete historical information can only be obtained from contract records or management records. The government analyst has a great advantage over his industry counterpart in this area since he has access to a broader data base composed of data collected from several contractors. The industry analyst may have only the historical data of his firm (Ref 21:12).

Data Base Adjustments. To ensure the data base is consistent and comparable several adjustments must be made to the raw data collected. Three instances where adjustments must be made to the data are for definitional differences, production quantity differences, and yearly price changes. Other types of adjustments, such as for contractor efficiency, for contract type, and for program stretch-out, are largely due to judgment and opinion. Research has not been done to treat these variables in a definitive manner; therefore, only the first three types will be discussed (Ref 21:17).

The first step in establishing a data base is to state the definitions and adjust the data to this definition.

For example, physical and performance type data has this problem of inconsistency of definitions. Speed data may be maximum speed or maximum speed at optimal altitude. Weight data can be maximum gross weight, takeoff weight or empty weight. Definitional differences are quite common due to the various accounting practices that are used in industry and make or buy arrangements. In many instances, the government has required the contractor to report cost to them by categories different from those used by the contractor's own internal system of accounting. This leads to inaccurate categorization of cost data, especially if costs are required to be broken down by recurring and nonrecurring categories. The problem of separating cost into recurring and nonrecurring is discussed in more detail later in this section since it plays an important role in the estimation of production costs from early actual data. Another reason for this problem is due to the government changing categories from time to time. To make data comparable over time, some programs require adjustments for definitional differences (Ref 21:18,21).

Cost-quantity adjustments are very important. If the cost-quantity relationship is ignored large errors will result. In most production processes, costs are generally a function of quantity, and cost is associated with a given quantity. In order to compare the costs of two systems when the quantities are the same, the costs can be adjusted with knowledge of the learning-curve slope. If costs are the same and the number of units are different, the comparison can be made by an adjustment to a common unit, e.g. 100th unit. Many times discounts are given for large purchases. Cost-quantity adjustments also must be considered in establishment of a consistent and comparable data base (Ref 21:30).

Price-level changes are the third type of adjustments that will be discussed. Price-level changes are generally due to inflation in addition to other factors which cause yearly price increases and devaluation of the dollar. This adjustment is made by a price index based on a particular base year applied to a dollar value of cost for any category. This problem is less severe and comparability improves when cost information can be collected in hours, e.g. labor hours, and a current rate applied to the hours to determine the cost in dollars. However, materials and purchase parts require the adjustment to be made. Some problems exist in identifying a base year for the indices and the year in which expenditures occurred. Identifying these factors can be difficult because in many cases the only data available is total contract cost (Ref 21:23,27).

Recurring and Nonrecurring Costs. For some decisions based on cost estimates, breakdown of functional cost categories into the categories of recurring and nonrecurring costs is desirable. As mentioned previously, decisions concerning follow-on production lots require the breakdown of costs into these categories. Recurring costs are costs that are incurred throughout the program and are a function of the number of aircraft produced. The nonrecurring costs are costs that are incurred once during the life of the program. Some examples of nonrecurring costs include preliminary design work, mockups, static tests, and initial startup costs (i.e. tooling and production planning, tool design, tooling manufacturer, purchase tooling and tool checkout (Ref 29:II-4)). Recurring costs are associated with a normal engineering requirements to keep a production system operating, normal equipment maintenance, and manufacturing labor and material used for producing airframes (Ref 17:8). Separation into these categories is very difficult if not impossible, especially for older aircraft models where the costs were not recorded using these categories (Ref 17 and 29:8 and II-15). Due to this limitation many experts in the field of cost analysis use judgment and experience from past studies to make the separation of these costs. Although not empirically possible to verify the reasonableness of this approach, the separation made must be evaluated on how logical it is and the acceptability of comparison of the results with past experience (Ref 29:II-15). However, according to the Large Study, "a determination of a breakdown into these categories after the fact tends to be questionable and introduces error into the data (Ref 17:8)."

Engineering and tooling labor are two functional categories that are difficult to separate into nonrecurring and recurring cost categories, and do not lend themselves to learning-curve analysis due to the non-repetitive nature of the work. Engineering is a function of time rather than the number of production units. For example, an engineering department may be working on a number of problems at one time which do not have a relationship to the number of units being produced. According to McDonald, "Tooling labor is usually a nonrecurring cost based on the time of tooling and the levels of production anticipated (Ref 20:I-10)." However, management has the decision to charge tooling costs directly to a production lot or amortize the cost over future anticipated produciton units (Ref 20:I-10). In the past, the government and industry management spent a great deal of time and effort to determine a basis for amortizing the initial startup costs. The contractors desired a short time period in which to recover their investments. The government auditors desired to defer portions of the expense to obtain a pro rata write-off over all the production benefitting from the initial investment. Many problems developed in specifying the timing of the amortization period, and in the case of tooling determining obsolescence (Ref 35:191). Trueger states in <u>Accounting Guide for Defense Contracts</u>.

> The Government has come to the realization that preproduction costs must be paid and is now, generally, quite willing to reimburse the contractor for such costs coincident with the first price redetermination so long as the armed services representatives are assured that no danger exists for possible duplication of such charges in subsequent periods (Ref 35:191).

Methods have been developed to separate costs into these separate components. Using the definitions in the CIR and the methods developed to separate costs into recurring and nonrecurring categories have made current data more consistent and comparable (Ref 21:21). According to the Air Force Systems Command Pamphlet, AFLCP/AFSCP 800-15, Acquisition Management Contractor Cost Data Reporting (Ref 1), the following are general principles that are applied to differentiate between nonrecurring and recurring cost categories:

I. Nonrecurring

a. Preliminary design effort encompassing the translation of weapon systems concpets and requirements into specifications for new systems as well as for major modification of existing systems.

b. Design engineering that entails the specifications and preparation of the original set of detailed drawings for new systems as well as for major modification of existing systems.

c. With respect to (a) and (b) above, it is preferable to identify the point of segregation between nonrecurring and recurring engineering costs as a specific event or point at which "design freeze" takes place as a result of a formal test or inspection, and after which formal engineering change proposal procedures must be followed to change design. If no reasonable event can be specified for this purpose, then all engineering drawing release may be used. The precise method used for segregating, recurring and nonrecurring engineering cost will be identified and explained in the "Remarks" space of the CIR.

d. System test and evaluation.

e. All partically completed reporting elements manufactured for tests (e.g. static, fatigue).

f. Costs of all tooling, manufacturing, and procurement effort specifically incurred in performing development or tests except for manufacture of complete units during the development program.

h. Training of service instructor personnel.

i. Initial preparation of technical data and manuals.

j. Start-up costs such as plant lay-out, operations, planning, plant rearrangement, tooling design and planning the original industrial engineering efforts to perfect a manufacturing technique.

II. Recurring

a. Engineering required for redesign, modification, reliability, maintainability, associated evaluation and liaison.

b. Complete reporting elements produced either for test (e.g. R&D flight test, operational evaluation, flight test, quality assurance, design evaluations, etc.), or for operational use.

c. Tool maintenance, modification, rework, and replacement.

d. Training all service personnel to operate and maintain equipment.

e. Reproduction and updating of technical data and manuals (Ref 1:4-5,4-6).

Data Base Used

The data base used in this thesis is cost data for nine fighter aircraft obtained from the data base used by

J. R. Large, H. G. Campbell and D. Cates for development of CERs in this report, Parametric Equations for Estimating Aircraft Airframe Costs (Ref 17). This report deletes all aircraft with first flight dates prior to 1953 from the available data base, hopefully, reducing variances due to technological advances in design, production systems, and cost data collection systems.

Copies of the Large study work sheets are on file in the Cost Library of the Aeronautical Systems Division located at Wright-Patterson AFB, Ohio. The work sheets provide a breakdown of hours and then-year dollars into both recurring and nonrecurring classifications of the various airframe cost accounting categories of engineering, tooling, manufacturing labor and material, quality control, and flight test.

Engineering and Tooling Categories. Engineering as defined by the Large study, "refers both to engineering for the basic airframe and to the system engineering performed by the prime contractor (Ref 17:18)." This category includes engineering hours for design studies, wind tunnel models, and other system and subsystem tests, and engineering hours for tooling and production planning. Excluded were those engineering hours not directly attributable to the aircraft system (Ref 17:18).

Tooling is limited to only tools designed for a particular program (i.e. assembly tools, dies, jigs, etc.).

Tooling hours include all effort spent for tool and production planning, set up, and maintenance of the machines, dies, and jigs (Ref 17:23).

Marcotte found that CERs for nonrecurring and recurring classication of engineering and tooling were very unreliable. However, when recurring and nonrecurring cost data were combined within each functional category a reliable CER which could predict the total cost of the functional category resulted.

> Our original intent was to estimate nonrecurring and recurring hours separately, but regression analyses of hours reported by contractors as "nonrecurring" indicated discrepancies in the data. Consequently, cumulative total engineering hours were plotted for each aircraft, and values were read off the curves at 25, 50, 100, and 200 aircraft. (Ref 17:18)

The same difficulties were observed by Large and associates in the tooling hour data. Therefore, only aggregated recurring and nonrecurring costs are used in this study to predict the costs of engineering and tooling for follow-on production. A better prediction of follow-on unit cost probably could be obtained if the recurring portion of total cost could be separated with a high degree of confidence.

Manufacturing Labor and Material Categories. Manufacturing labor and material are much easier to separate into the components of recurring and nonrecurring cost categories. This is due to clearer definitions and more uniform accounting procedures for dealing with these functional cost categories.

In the Large study work sheets, manufacturing labor is accounted for in hours, therefore, a price index is not required for comparative adjustments. Usually, a standard wage rate per hour is applied to obtain a dollar figure. This category includes all direct labor hours for assembling the major structure of the airframe and installation of purchase parts and equipment, whether contractor or government furnished.

Manufacturing materials include the raw and semifabricated materials. Also included are purchased equipment, standard hardware items (i.e. fittings and rivets, government furnished items [radios and avionics], and contractor procured items (Ref 17:31)). The manufacturing material costs provided in the Large study work sheets are broken down into recurring and nonrecurring costs of the components of materials, purchased equipment, and Government Furhished Aerospace Equipment (GFAE). These costs are in then-year dollars; requiring a conversion to constant 1973 year dollars to improve comparability.

The conversion to constant year dollars was accomplished using aircraft acceptance schedules provided by J. Large. The year a specific airframe was accepted by the government was assumed to be the year in which the cost was incurred. Since material costs were aggregated by lot in the CIR and the Large study work sheets, the costs were broken out by year and the appropriate price index applied. The breakout was accomplished by summing the number of airframes accepted in one year from the acceptance schedules and forming a ratio to total airframes in the lot. Multiplying this ratio times the cumulative cost of the total lot of airframes resulted in the cumulative cost of airframes by year. Once the cost was broken down by year, the cost per year was multiplied by the price indices provided in the Large study to convert dollar cost to constant 1973 dollars (Ref 17:32). Purchased equipment and GFAE were aggregated and multipled by the price index for purchased equipment resulting in a total purchased equipment cost. After this correction was performed, the constant 1973 dollar material cost and total purchase equipment cost were aggregated by the original lot quantities into a total material cost. This methodology was employed to obtain material cost for each airframe in the sample selected for the study except for the F-14 airframe program.

The acceptance schedule for the F-14 airframe program was not available, therefore, the schedule was estimated. The F-14 data work sheet indicated five lots of F-14 airframes were accepted. The estimated acceptance schedule was based on the assumption that a lot of F-14 airframes was accepted each year beginning with the year that the first flight occurred, 1970. Since 1973 was the base year used for the Large study, the last two lots of F-14 airframes were assumed to have been accepted in 1973.

Other Categories. Cost data for Aerospace Ground Equipment (AGE), Training, Spares, General and Administrative, Quality Control, and Flight Test expenses are also provided on the work sheets. These costs were not used for any of the analyses performed in this study.

Physical and Performance Data Source. The physical and performance data was obtained from two sources. The speeds, maximum at best altitude, of the various airframe programs used in the sample were obtained from the Large study (Ref 17:13), and consisted of the value of speed for the one hundredth airframe. This value was assumed constant for all lots of each type airframe produced. The value of speed for the F-14 was a recent estimate released to the public. The weight data was obtained from the data used by V. Handel in his thesis, Aircraft Airframe Cost Estimation by the Application of Joint Generalized Least-Squares, (Ref 13). The weights fluctuate between the different lots due to modifications of the airframe. The value of weight used for the analyses was the weight for the lot in which the airframe quantity being analyzed fell. For example, if a first lot of airframes, airframes 1 to 20, weighed 1200 pounds and the second lot, airframes 21 to 72, weighed 1400 pounds, the 50th airframe was assumed to weigh 1400 pounds.

Because the data is privileged information, it is included in Appendix A. The Appendix shows cumulative cost data by lot, broken down into engineering hours, tooling hours, recurring labor hours, and recurring material cost in 1973 dollars. If one, two, or three airframes were produced in the first lot, the recurring costs for that lot were discarded and assumed to be non-recurring costs associated with a prototype effort. When this assumption was made, the number of units in the first lot were included in the cumulative quantity because some learning has resulted from work on these first units.

Development of Marginal CER using Large Methodology

The marginal or unit average cost CERs developed by Marcotte and the Bayesian approach provide estimates of cost using quantity as one of the explanatory variables in addition to weight and speed. The Large study developed equations which estimate the cost of a specific airframe unit in terms of the explanatory variables of weight and speed. The methodology employed by Large represents the current state-of-the-art and was selected for comparing the predictive capability of the Bayesian approach. In order to compare estimates using the Bayesian equations with estimates using equations with weight and speed as the only explanatory variables, marginal CERs must be developed using the methodology employed by Large and the data base of fighter airframes.

Development of marginal relationships involve calculation of an average cost per lot (C*) which is obtained by dividing cumulative lot cost by the quantity of units in the lot. A problem arises due to the learning-curve assumptions applicable to CER development. The problem is where along the quantity axis should the average cost per lot be plotted. Since the log-linear function of learning-curve theory is applicable, the use of arithmetic midpoints produces an unequal distribution of area under the curve, figure 5. The lower graph in figure 5 illustrates the rational for using a true lot midpoint.

The true lot midpoint is the unit (Q*) which represents the entire lot and reflects the average cost (C*) of that lot. The total cost of the lot is equal to the product of C* times Q*. The product approximates the area under the curve for the number of the units in the lot, illustrated by the shaded areas of approximately equal size on the lower graph of figure 5 (Ref 21:105).

The first step in the development of a marginal CER for estimating the cost of a functional category of cost in terms of weight and speed was performance of regression analysis on the data of each airframe program in the data base. The dependent variable of cumulative cost per lot and the independent variable of cumulative quantity by lot were used for the regression analysis. The resulting CERs estimate cumulative functional cost directly from cumulative quantity for each airframe program in the data base. Table I presents the CERs for engineering hours (E_H),

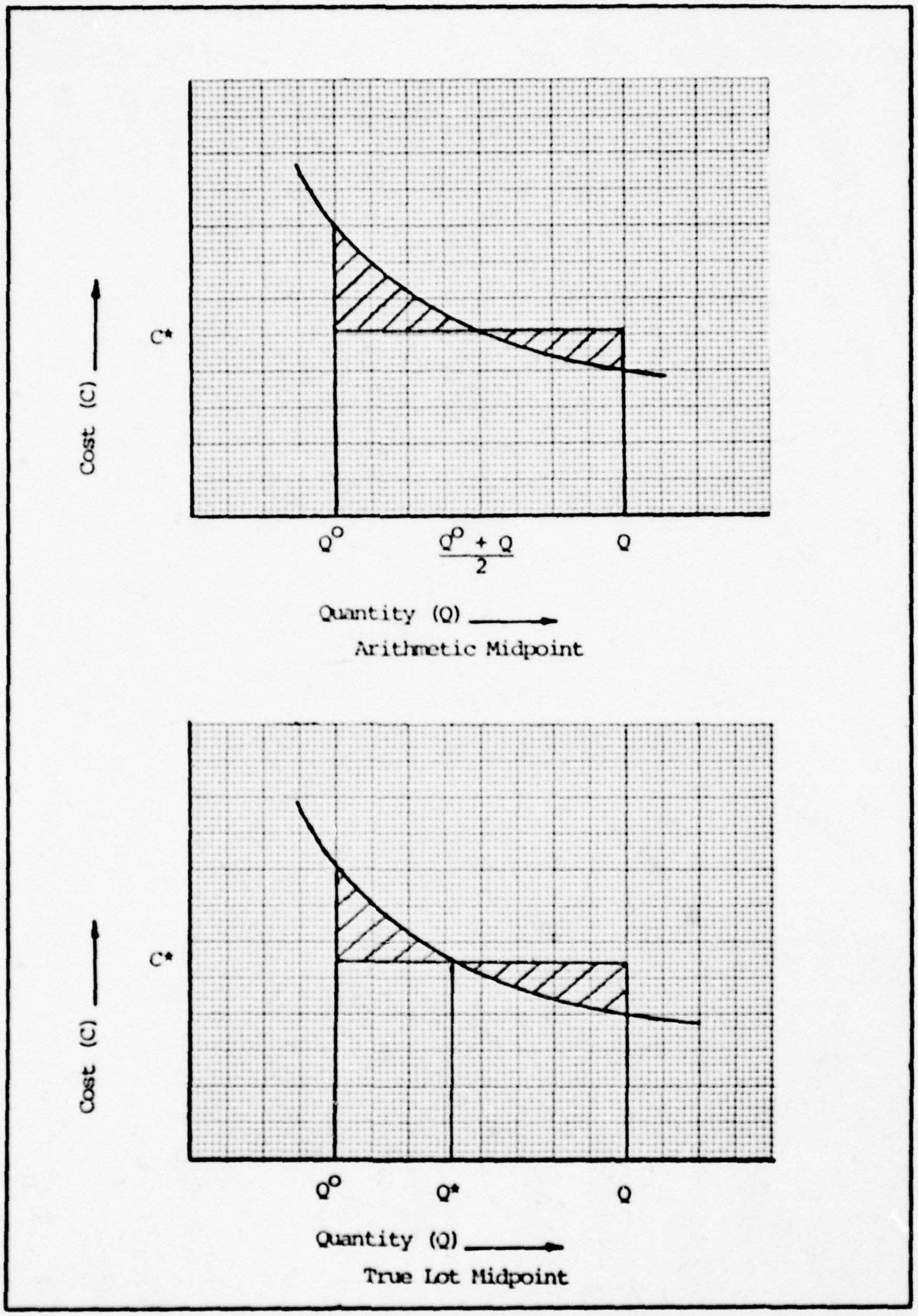


Figure 5. Comparison of True Lot with Arithmetic Midpoint (Ref 21:105)

Table I
CER to Predict Cost from Cumulative Quantity
and
Learning Curve Slopes Derived from Quantity Coefficient

Type A/C	CER	Learning Curve Slope
A-4D	$E_H = e^{13.3861} Q^{.1691}$	56%
	$T_H = e^{13.0175} Q^{.4683}$	69%
	$I_H = e^{12.9390} Q^{.5769}$	75%
	$M_C = e^{13.3310} Q^{.8606}$	91%
F-102	$E_H = e^{14.5636} Q^{.1896}$	57%
	$T_H = e^{15.0570} Q^{.2323}$	59%
	$L_H = e^{13.6853} Q^{.6041}$	76%
	$M_C = e^{13.4405} Q^{.9218}$	95%
F-104	$E_H = e^{13.6917} Q^{.3521}$	64%
	$T_H = e^{13.5257} Q^{.3794}$	65%
	$L_H = e^{12.1800} Q^{.7215}$	82%
	$M_C = e^{13.0882} Q^{.9037}$	94%
F-105	$E_H = e^{15.1139} Q^{.2498}$	59%
	$T_H = e^{13.953} Q^{.3768}$	65%
	$L_H = e^{14.4209} Q^{.5632}$	74%
	$M_C = e^{14.4908} Q^{.8716}$	91%

Table I (continued)

Type A/C	CER	Learning Curve Slope
F-106	$E_H = e^{14.3660} Q^{.2585}$	60%
	$T_H = e^{14.6089} Q^{.2856}$	61%
	$L_H = e^{13.1255} Q^{.6957}$	81%
	$M_C = e^{15.5919} Q^{.6725}$	80%
A-5	$E_H = e^{15.2581} Q^{.2509}$	60%
	$T_H = e^{14.4362} Q^{.2778}$	61%
	$L_H = e^{13.4611} Q^{.7315}$	83%
	$M_C = e^{15.5079} Q^{.7244}$	83%
F-4	$E_H = e^{15.3933} Q^{.1737}$	56%
	$T_H = e^{15.3460} Q^{.1301}$	55%
	$L_H = e^{14.0555} Q^{.6239}$	77%
	$M_C = e^{15.4487} Q^{.7246}$	83%
F-111	$E_H = e^{16.3241} Q^{.1370}$	55%
	$T_H = e^{14.8401} Q^{.3685}$	65%
	$L_H = e^{13.1034} Q^{.8533}$	90%
	$M_C = e^{15.7443} Q^{.8001}$	87%

Table I (continued)

Type A/C	CER	Learning Curve Slope
T-38	$E_H = e^{12.1339}\ Q^{.4705}$	69%
	$T_H = e^{13.0546}\ Q^{.3672}$	64%
	$L_H = e^{11.5568}\ Q^{.7371}$	83%
	$M_C = e^{13.1324}\ Q^{.9263}$	95%

tooling hours (T_H), labor hours (L_H), and material costs in dollars (M_C) developed from the first step.

The second step required development of a marginal CER in terms of quantity for each functional cost category of each airframe program. The marginal CERs were obtained from regression of average cost per lot by functional cost. The true lot midpoints were determined using tables of learning-curves published by the RAND Corporation (Ref 7) and the equations in Table I. The coefficient of the variable quantity of the cumulative CER was used to calculate the slope of the unit learning-curve using the following relationship,

$$Y_C = aX^B \tag{13}$$

where,

Y_C = cumulative average cost

a = constant

X = variable quantity

b = coefficient of X, equal to ln S/ln 2

s = slope of the learning-curve

and

$T = Y_c X$, where t equals cumulative cost of X units

$$T = aX^b X \tag{14}$$

$$= aX^{b + 1} \tag{15}$$

and, k = b + 1, coefficient of cumulative quantity term of the cumulative cost CER

$$k = \frac{\text{Ln } S}{\text{Ln } 2} + 1 \tag{16}$$

Once the unit learning-curve slope(s) is known the tables of *learning-curve* slopes were consulted to obtain the true lot midpoints for the lot quantities of the data base.

The last step is to obtain the marginal CER in terms of the explanatory variables of weight and speed. Using the marginal CERs developed in step two, an estimate of unit cost for the 50th and 100th airframe was calculated by functional cost category and type aircraft.

The estimates and step two marginal CERs are shown in Tables II-V. The values of cost by functional category for the 50th unit were regressed against the values for speed and weight, Table VI, to obtain the marginal CERs

Table II
CERs to Calculate Engineering Hours
for a Given Marginal Airframe Unit

Type Aircraft	CER	Engineering Hours	
		Q = 50	Q = 100
A-4D	$E_H = e^{13.7400} Q^{-1.2503}$	6,966	2,928
F-102	$E_H = e^{13.120} Q^{-.8023}$	21,617	12,396
F-104	$E_H = e^{13.0598} Q^{-.6184}$	41,792	27,223
F-105	$E_H = e^{14.4660} Q^{-.8993}$	56,845	30,478
F-106	$E_H = e^{14.3400} Q^{-1.0234}$	30,839	15,171
A-5	$E_H = e^{14.1067} Q^{-.7309}$	76,672	46,196
F-4	$E_H = e^{14.3606} Q^{-.9800}$	37,298	18,909
F-111	$E_H = e^{15.5429} Q^{-1.0982}$	76,632	35,795
T-38	$E_H = e^{11.8200} Q^{-.6270}$	11,699	7,575

Table III
CERs to Calculate Tooling Hours
for a Given Marginal Airframe Unit

Type Aircraft	CER	Tooling Hours	
		Q = 50	Q = 100
A-4D	$T_H = e^{12.1776} Q^{-.6954}$	12,798	7,903
F-102	$T_H = e^{14.1670} Q^{-.8311}$	55,043	30,941
F-104	$T_H = e^{13.5274} Q^{-.8021}$	32,521	18,652
F-105	$T_H = e^{13.7241} Q^{-.7990}$	40,074	23,033
F-106	$T_H = e^{14.8065} Q^{-.9905}$	55,919	28,145
A-5	$T_H = e^{13.5541} Q^{-.7961}$	34,196	19,694
F-4	$T_H = e^{14.0085} Q^{-1.0114}$	23,198	11,508
F-111	$T_H = e^{14.4103} Q^{-.7715}$	88,619	51,913
T-38	$T_H = e^{12.8361} Q^{-.8579}$	13,092	7,224

Table IV
CERs to Calculate Recurring Labor Hours
for a Given Marginal Airframe Unit

Type Aircraft	CER	Labor Hours	
		Q = 50	Q = 100
A-4D	$L_H = e^{12.6762} Q^{-.4949}$	46,167	32,760
F-102	$L_H = e^{13.1106} Q^{-.3598}$	120,928	94,234
F-104	$L_H = e^{12.3064} Q^{-.3638}$	53,277	41,402
F-105	$L_H = e^{14.0175} Q^{-.4620}$	200,846	145,814
F-106	$L_H = e^{13.1278} Q^{-.3564}$	124,679	97,386
A-5	$L_H = e^{13.0876} Q^{-.2353}$	192,375	163,429
F-4	$L_H = e^{13.8485} Q^{-.4463}$	180,305	132,327
F-111	$L_H = e^{13.0176} Q^{-.1674}$	233,902	208,277
T-38	$L_H = e^{11.4930} Q^{-.3232}$	27,681	22,124

Table V
CERs to Calculate Recurring Material Cost
for a Given Marginal Airframe Unit

Type Aircraft	CER	Material Cost 1973 $	
		Q = 50	Q = 100
A-4D	$M_C = e^{13.4601}\ Q^{-.2160}$	301,051	259,187
F-102	$M_C = e^{13.4952}\ Q^{-.1030}$	485,116	451,675
F-104	$M_C = e^{13.1341}\ Q^{-.0957}$	347,968	325,644
F-105	$M_C = e^{14.0751}\ Q^{-.1322}$	772,917	705,243
F-106	$M_C = e^{15.0668}\ Q^{-.3041}$	1,063,454	861,315
A-5	$M_C = e^{15.1435}\ Q^{-.2477}$	1,431,758	1,205,875
F-4	$M_C = e^{15.2426}\ Q^{-.3012}$	1,282,534	1,040,882
F-111	$M_C = e^{15.5183}\ Q^{-.2002}$	2,584,859	2,249,955
T-38	$M_C = e^{13.1765}\ Q^{-.1050}$	349,946	325,370

Table VI
Data Used to Calculate CERs for Predicting Airframe Marginal Cost Estimates for the 50th and 100th Unit

Type Aircraft	Engineering Hours	Tooling Hours	Recurring Labor Hours	Recurring Material 1973 $	Weight Pounds	Speed Knots
Q = 50						
A-4D	6,966	12,798	46,167	301,051	5,162	578
F-102	21,617	55,043	120,928	485,116	12,052	677
F-104	41,792	32,521	53,277	347,968	8,069	1,150
F-105	56,845	40,074	200,846	772,917	19,758	1,195
F-106	30,839	55,919	124,679	1,063,454	14,630	1,153
A-5	76,672	34,196	192,375	1,431,758	23,828	1,147
F-4	37,298	23,198	180,305	1,282,534	17,320	1,220
F-111	76,632	88,619	233,902	2,584,859	32,926	1,262
T-38	11,699	13,092	27,681	349,946	5,348	750
Q = 100						
A-4D	2,928	7,903	32,760	259,187	5,072	578
F-102	12,396	30,941	94.234	451,675	12,052	677
F-104	27,223	18,652	41,402	325,644	8,069	1,150
F-105	30,478	23,033	145,814	705,243	18,911	1,195
F-106	15,171	28,145	97,386	861,315	15,400	1,153
A-5	46,196	19,694	163,429	1,205,875	26,613	1,147
F-4	18,909	11,508	132,327	1,040,882	18,666	1,220
F-111	35,795	51,913	208,277	2,249,955	32,926	1,262
T-38	7,575	7,224	22,124	325,370	5,482	750

in terms of these explanatory variables. The same methodology was employed to obtain CERs for the 100th airframe. Table VII contains the marginal CERs developed which will be used to provide theestimates for comparison in a later chapter.

Application of Marginal CERs. The slope of the learning-curve caluclated from the equations for 50th and 100th unit is assumed to hold throughout production. This slope is then applied to the 100th unit to determine the cost of the first unit. The first unit cost and slope are used to determine the cost of the X unit using the following equations, (Ref 17:55).

$$\frac{\hat{Y}_{100}}{\hat{Y}_{50}} = \text{slope (S)} \tag{17}$$

$$b = \frac{\log S}{\log 2} \tag{18}$$

$$\hat{Y}_{100} = \hat{Y}_1 (100)^b \tag{19}$$

$$\hat{Y}_1 = \frac{\hat{Y}_x}{(x)^b} \tag{20}$$

$$\hat{Y}_x = \hat{Y}_1 (x)^b \tag{21}$$

where,

$\hat{Y}_i$ = estimate of cost of i unit

S = slope of learning-curve

Table VII

CERs Developed from RAND Data to Predict Marginal Unit Cost for the 50th and 100th Unit from Knowledge of Weight and Speed*

CER							R^2
E_{50} =	e	-12.3484	W	.6443	S	1.4068	.95
		(-5.18)		(2.70)		(2.72)	
E_{100} =	e	-13.4565	W	.5042	S	1.6691	.92
		(-3.84)		(1.47)		(2.18)	
T_{50} =	e	-3.2196	W	.8514	S	-.1955	.97
		(-.82)		(2.17)		(-.23)	
T_{100} =	e	-2.9404	W	.8162	S	-.2748	.96
		(-.70)		(1.98)		(-.30)	
L_{50} =	e	-4.6180	W	1.2660	S	-.3937	.98
		(-2.38)		(6.53)		(-.94)	
L_{100} =	e	-4.9727	W	1.3037	S	-.4342	.97
		(-3.27)		(8.74)		(-1.31)	
M_{50} =	e	-4.2081	W	1.0351	S	.1474	.99
		(-1.66)		(4.08)		(.27)	
M_{100} =	e	-3.5810	W	1.0149	S	.0617	.99
		(-1.74)		(5.01)		(.14)	

*t-statistic given in parentheses.

b = coefficient of variable quantity

x = particular unit for which an estimate of cost is desired

Summary

Some general information concerning data bases was discussed in this chapter. The information included the types of data required for development of CERs and some adjustments that are made to data to ensure consistency and comparability. This discussion was followed by a detailed discussion of the definitional differences and difficulties associated with separating cost into non-recurring and recurring cost classifications. The next section enumerated and provided the explanation of the procedure used to convert the work sheet then-year dollars data to constant year dollars for the analyses of the material category. The last section provided an explanation of the methodology employed to obtain marginal CERs having explanatory variables of weight and speed for estimating total engineering and tooling hours, recurring labor hours and material dollars.

IV. Methodology

The methodology employed in this thesis integrates two concepts: generalized least-squares (GLS) and a Bayesian approach to multiple regression. GLS is a methodology that allows ordinary least-squares to be used to estimate coefficients and variances for a general linear model having correlated error terms. The concept of a Bayesian approach to multiple regression allows the use of available information, both historical and current information from a new experiment, to update estimates of the coefficients of a linear regression equation, CER. The application of a Bayesian approach to the problem of predicting future lot costs should have particular appeal to the analyst who performs cost estimation of airframe programs assumed to have similar design and production variances found in past programs.

The RANDOM CERs developed by Marcotte (Ref 19) from historical data are used for the application of the Bayesian approach in this research effort because the RANDOM CERs were proven to be statistically "better" estimators of cost than CERs derived using other recently published methodologies. Since the RANDOM CERs were developed from a model which considered correlation, prediction using the Bayesian updated RANDOM CERs poses a problem that requires special treatment in the GLS model.

The first section of this chapter provides a brief introduction to Bayesian philosophy and includes an example of a Bayesian application to a discrete case. The next two sections provide discussions of the two concepts employed in this research. The final section discusses the application of GLS and a Bayesian approach to multiple regression for updating an airframe CER to predict the cost of the next unit. Also included in this section is a brief discussion of how the 95 percent prediction interval was calculated.

Bayesian Philosophy

The Bayesian approach is concerned with predicting the occurrence of some underlying "state of nature" that is uncertian. The Bayesian approach uses the available evidence to assess a risk to each of the various alternative actions for solving a problem (Ref 23:1). Some problems involve the implementation of various decisions. These problems fall in the realm of statistical decision theory. Other problems require only one action, either to accept or reject a hypothesis. This is referred to as statistical inference. The problems involving estimation are of this type. The decision rule is the estimator and the action is the estimate. The Bayesian estimator refers to a decision rule with the smallest expected risk (Ref 30:358).

The development of the theory was made by an English Clergyman, Reverend Thomas Bayes, during the 18th century (Ref 16:314). Since that time Bayesian philosophy has found application to many problems involving inferences from prior information. Bayesian philosophy has several interesting aspects.

Aspects. The first aspect is a basic concept which allows the use of prior information about probabilities or evidence. This feature of the explicit treatment of prior evidence is distinctive of Bayesian statistics. The prior information in classical statistics is only considered informally, if at all. The Bayes' Theorem uses both the prior information and additional evidence gathered (i.e. sampling) to revise the prior information. The vaule of the Bayesian approach is only to the extent that both types of information are related to the uncertain future state of nature. As more information is gathered, less uncertainty exists. The Bayesian analysis for updating the prior information is only as good as the additional information obtained. If the additional information gathered is related to the future state of nature, more information results in less uncertainty (Ref 22:2,3).

The main aspect of this philosophy is that prior probabilities can be either subjective or objectively determined. The objective probabilities for example may be obtained from historical data which is the case in this research. However, the greatest power of Bayesian analysis can be derived when no prior information is available and subjective probabilities are used. Many times the subjective probabilities are nothing more than "gut feelings" that evolve from experience and judgment (Ref 16:318).

The use of subjective information from intuition, judgment, and feelings directly in the formal analysis of a decision problem is the primary philosphical difference between the Bayesian statistician and the "classical" statistician. The Bayesian statistician tries to make use of judgments and experience, and infers that these count for something (Ref 22:3). The "Classical" approach believes that subjective aspects should be left out of the formal analysis (Ref 28:XX). These probabilities, whether determined subjectively or objectively, are referred to as "a priori" probabilities. The probabilities obtained from observation and application of Bayes' Theorem are referred to as "a posterior" probabilities (Ref 23: 439).

Another aspect is the philosophical differences in interpretation. The results obtained by both Bayesian and classical approaches are the same, but the interpretation is subject to differences that are philosophical rather than methodological. "Classical" statistics basically start with the premise that some true value of the parameter exists to be estimated, for example a mean.

AD-A077 064 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOO--ETC F/G 15/5
APPLICATION OF A BAYESIAN APPROACH TO UPDATING AIRFRAME CERS.(U)
SEP 79 W D DIETRICH
UNCLASSIFIED AFIT/GSM/SM/76D-30 NL

2 OF 2
AD-
A077064

END
DATE
FILMED
12-79
DDC

Using a sample, a confidence interval is developed. The confidence interval is established with some probability that the true value of the parameter falls within some specified interval (i.e. ninety-five percent probability the confidence interval contains the true value of the mean). The estimate of the true value of the parameter is the random variable.

In Bayesian statistics the parameter is treated as the random variable. The existance of a true value of the parameter is seen as irrelevant since the true value cannot be observed. This treatment of the parameter as a random variable is particularly of interest for predicting future events. Bayesian inference assigns probabilities directly to the "parameter space" which is the range of values the parameter can take on (Ref 22:48).

Another major aspect of philosophical difference is that the underlying probability distribution of the random variable of the prior distribution in Bayesian analysis can be assumed to be any form, and can be based on judgment or past objective data. The two distribution functions that are easiest to deal with mathematically are the diffuse or uniform and the normal or Gaussian (Ref 22:45).

Discrete Application of Bayes' Rule. As discussed in the previous paragraphs of this section, Bayes' Rule (Theorem) provides a mathematical formula by which probabilities of prior events can be updated given additional information. If the probability and corresponding value of the payoff are known, application of Bayes' Rule can result in the expected value or an updated estimate of cost. The Bayesian probability revisions provide the decision maker with a measure of the effect of the additional information (Ref 27:110:

> If B_1, $B_2, \ldots,$ and B_k constitute a set of mutually exclusive events of which one must occur and none has a zero probability, then for any event A for which $P(A) \neq 0$
>
> $$P(B_r|A) = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)} \qquad (22)$$
>
> for $r = 1, 2, \ldots,$ or k (Ref 11:62).

where,

$P(B_r)$ = the prior probability of B_r

A = the observation

$P(A|B_r)$ = the probability of the observation given B_r is true

$P(B_r|A)$ = the posterior probability of B_r

The denominator of this expression is simply P(A) from a theorem called the rule of elimination. This theorem states:

> If B_1, $B_2, \ldots,$ and B_k constitute mutually exclusive events of which one must occur and none has a zero probability, then for any event A (Ref 11:59).

$$P(A) = \sum_{i=1}^{K} P(B_i) \cdot P(A|B_i) \qquad (23)$$

A simple problem consisting of three actions with one intermediate state is presented to illustrate Bayes' Rule. The following paragraph is a statement of the problem.

In a tactical war game, the enemy can take one and only one action as their next move in the game. The first action denoted as B_1 is to attack using tactical nuclear weapons. The second action denoted B_2 is to attack using conventional weapons. The last action denoted B_3 is to retreat. Our intelligence section has determined either through experience or expert judgment the following probabilities for each action, where $P(B_n)$ is the probability of the n^{th} action.

$P(B_1) = .3$

$P(B_2) = .6$

$P(B_3) = .1$

The intelligence experts have also determined that there is a .2 probability that the enemy will precede a nuclear attack with a reconnaissance flight, a .7 probability that they will precede a conventional attack with a reconnaissance flight, and a .1 probability that a retreat will be preceded by a reconnaissance flight. As the staff is discussing various strategies for each of the possible enemy actions, an enemy photo reconnaissance plane is sighted.

With this additional information, what are the probabilities for each of the three possible enemy actions.

The following is the general expression for the above problem:

$$P(B_n|RF) = \frac{P(B_n) \cdot P(RF|B_n)}{\sum_{i=1}^{k} P(B_i) \cdot P(RF|B_i)} \tag{24}$$

where,

$P(B_n)$ = probability of one of the n actions

$P(B_n|RF)$ = probability of one of the n actions given that a reconnaissance flight has been observed. This is referred to as the posterior probability.

$P(RF|B_n)$ = probability of a reconnaissance flight given one of n actions occurs. This probability is referred to as the likelihood.

Using the above notation to formulate, the revised probability of a nuclear attack given that a reconnaissance flight has been sighted is as follows:

$$P(B_1|RF) = \frac{P(B_1) \cdot P(RF|B_1)}{P(B_1) \cdot P(RF|B_1) + P(B_2) \cdot P(RF|B_2) + P(B_3) \cdot P(RF|B_3)}$$

(4)

The specific values of the posterior probabilities of the three possible actions are indicated as follows:

$$P(B_1 | RF) = \frac{0.3 \times 0.2}{0.3 \times 0.2 + 0.6 \times 0.7 + 0.1 \times 0.1} = .123$$

$$P(B_2 | RF) = \frac{0.6 \times 0.7}{0.3 \times 0.2 + 0.6 \times 0.7 + 0.1 \times 0.1} = .857$$

$$P(B_3 | RF) = \frac{0.1 \times 0.1}{0.3 \times 0.2 + 0.6 \times 0.7 + 0.1 \times 0.1} = .02$$

Summary. Decision makers desire to decrease the uncertainty that exists in the decision required to be made in this complex world. Modern decision makers spend much time and money gathering additional information to improve their knowledge of the future outcomes or events (Ref 16.3). Bayesian analysis provides a means for treating the additional information with probabilities determined from prior information which results in a revised probability of a future event occurring that is maximum likelihood with minimum variance. The Bayesian approach also has an intuitive appeal to the decision maker because this approach allows the subjective probabilities determined from judgment and experience to be used. For the estimator, the historical or prior information helps reduce complete ignorance of the cost of an item that is similar to items for which prior information is available.

Concept of Generalized Least-Squares (AITKEN) Estimators

Generalized Least-Squares (GLS) is a methodology which deals with linear models where correlation is assumed to exist among the disturbance terms. The methodology to obtain estimates of the regression coefficients and variance is presented in this section. To begin the presentation of the methodology the following paragraphs provide a comparison of the GLS and ordinary least-squares assumptions.

Assumptions of GLS and Ordinary Least-Squares. A linear model assumed to have correlated disturbance terms with a mean of zero requires the application of GLS to provide unbiased estimates of the parameters B, Var (B), and σ^2. Expressed mathematically, the model is of the form

$$Y = XB + \varepsilon \tag{25}$$

and the assumption

$$E(\varepsilon'\varepsilon) = \sigma^2 Q \tag{26}$$

where σ^2 is an unknown constant term and Q is a known symmetric positive definite matrix of n X n terms. Equation (26) infers that the error terms are correlated and that the variance and covariance of ε terms are known up to a scale factor (Ref 14:208).

The assumption expressed by equation (26) is significantly different from the assumption of ordinary least-squares. The assumptions of ordinary least-squares state,

1. The mean of the error term is equal to zero

$$E(\varepsilon) = 0 \tag{27}$$

2. The error terms are independent and the variance constant

$$E(\varepsilon'\varepsilon) = \sigma^2 I \tag{28}$$

where I represents an identify matrix with zeros in the off diagonal elements and σ^2 is constant (Ref 14:122).

$$E(\varepsilon'\varepsilon) = \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{29}$$

Positive Definite Matrix. A positive definite matrix is a matrix where every principle minor is positive. A simple example in Johnston (Ref 14) uses a 2 X 2 matrix for illustration. If Q is this 2 X 2 matrix, then

$$E(\varepsilon\varepsilon') = \sigma^2 Q = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \tag{30}$$

where σ_{ij} is the variance of ε_i and σ_{ij} is the covariance between ε_i and ε_j. To be positive definite $\sigma_{11}>0$, and $\sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1 - p_{12}^2)>0$ where p_{12} is the population correlation between ε_1 and ε_2. Translation of the previous mathematical statements is each error term must have a positive variance and the two errors are not perfectly correlated. For multiple dimensional matrices there must be no perfect correlation between pairs of error terms or between the set of all ϵ's (Ref 14:209).

Methodology to Estimate Parameters. Assuming the error terms are correlated the first problem is to estimate B of the linear model, $Y = XB + \epsilon$. This estimation can be approached in several ways; the presentation outlined in Johnston is the simplest (Ref 14:209).

A positive definite matrix (Q) can be expressed by PP', where P is nonsingular, therefore

$$Q = PP' \tag{31}$$

then

$$P^{-1}QP^{-1'} = I \tag{32}$$

and

$$P^{-1'}P^{-1} = Q^{-1} \tag{33}$$

Now the linear model is premultiplied by P^{-1} to give

$$P^{-1}Y = P^{-1}XB + P^{-1}\epsilon \tag{34}$$

where,

$$Y_* = P^{-1}Y, \; X_* = P^{-1}X, \text{ and } \epsilon_* = P^{-1}\epsilon$$

It can easily be seen using equation (32) that

$$E(\epsilon_*\epsilon_*') = \sigma^2 I \tag{35}$$

therefore, equation (35) satisfies the assumptions of ordinary least-squares. By performing ordinary least-squares on equation (34) the estimate of B is,

$$\hat{B} = (X_*{}' X_*)^{-1} X_*{}' Y_* \tag{36}$$

Making substitutions the *GLS* $\hat{B}$ estimate is

$$\hat{B} = (X' Q^{-1} X)^{-1} X' Q^{-1} Y \tag{37}$$

This estimate is a minimum variance unbiased estimator with the following variance-covariance matrix

$$\mathrm{VAR}(\hat{B}) = \sigma^2 (X' Q^{-1} X)^{-1} \tag{38}$$

The unbiased estimate for σ^2 using ordinary least-squares is

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}' \hat{\varepsilon}}{n-K} \tag{39}$$

where $\hat{\varepsilon} = Y - X\hat{B}$

For the general least-square model, an unbiased estimate for σ^2 is

$$\sigma^2 = \frac{1}{n-k} (Y_* - X_*\hat{B})' (Y_* - X_*\hat{B})$$

$$= \frac{1}{n-k} (Y - X\hat{B})' Q^{-1} (Y - X\hat{B}) \tag{40}$$

where n is the number of observations and k is the number of explanatory variables in the linear model (Ref 14: 210). The resultant difference of n less k is the number of degrees of freedom.

The unbiased estimate ($\hat{B}$) provided by the above methodology is a maximum likelihood estimator with minimum variance in the class of all unbiased estimators. The $\hat{B}$ estimator defined in equation (37) is the generalized least-squares (Aitken) estimator. Furthermore, if Q is know, the GLS estimator $\hat{B}$ can be computed using equation (37) and the standard error from equation (38). With this information, the tests for significance and confidence intervals can be calculated. For a more detailed explanation see Johnston (Ref 14:208-210).

Treatment of Prediction in the GLS Model. Johnston presents the following treatment of the prediction problem in the GLS model (Ref 14:212)

$$y = XB + \epsilon \tag{41}$$

with $E(\epsilon) = 0$ and $E(\epsilon\epsilon') = V$, which is assumed to be a known, symmetric, possitive-definite matrix. The problem is to predict the dependent variable y_0 given a set of independent variables X_0. The prediction equation becomes

$$y_0 = X_0B + \epsilon_0 \tag{42}$$

where ϵ_0 is the true, but unknown value of the disturbance. Assuming

$$E(\varepsilon_0) = 0 \tag{43}$$

$$E(\varepsilon_0^{\,2}) = \sigma_0^{\,2} \tag{44}$$

$$E(\varepsilon_0 \varepsilon) = \begin{bmatrix} E(\varepsilon_1 \varepsilon_0) \\ E(\varepsilon_2 \varepsilon_0) \\ \cdot \\ \cdot \\ \cdot \\ E(\varepsilon_n \varepsilon_0) \end{bmatrix} = W \tag{45}$$

where W is an n X 1 vector of convariances of the prediction disturbance with the vector of sample disturbances.

$$p = C'Y \tag{46}$$

such that C is a vector of n constants. If p is to be a best linear unbiased predictor, then a value of C must be determined that minimizes the prediction variance

$$\sigma_p^{\,2} = E\left\{ (p - Y_0)^2 \right\} \tag{47}$$

subject to $E\,(p - y_0) = 0$. C must satisfy the following expression for p to be an unbiased predictor.

$$C'X - X_0 = 0 \tag{48}$$

The prediction error is

$$p - y_0 = C'\varepsilon - \varepsilon_0 \tag{49}$$

which results in the prediction variance of

$$\sigma_p^2 = C'VC + \sigma_0^2 - 2C'W \tag{50}$$

Using LaGrange multipliers and differentiation, Johnston derives an estimate $\hat{C}$ such that

$$\hat{C} = V^{-1}[I - X\,(X'V^{-1}X)^{-1}X'V^{-1}]W + V^{-1}X(X'V^{-1}X_0{}^{-1})X_0' \tag{51}$$

Substituting the estimate of $\hat{C}'$ into equation (46), the following prediction equation results,

$$\grave{p} = X_0\hat{B} - W'V^{-1}\varepsilon \tag{52}$$

where $\varepsilon = Y - X\hat{B}$ is a vector of the GLS residuals and $\hat{X}_0B$ is the prediction not adjusted for the correlation existing among the observations. The product $W'V^{-1}\varepsilon$ is the adjustment factor which takes into account the correlation between the actual observations of the sample program and the prediction (Ref 14:212,213).

The following section presents the methodology for the application of a Bayesian approach to multiple regression. Using Bayesian methodology, the $\hat{B}$ and σ^2 estimates can be updated as new information is acquired.

A Bayesian Approach to Multiple Regression Analysis

The application of the Bayesian approach to multiple regression analysis is complex mathematically when compared with the discrete example presented in a preceding section of this chapter. However, the underlying concept of updating prior information is the same.

The following is the presentation of the application of a Bayesian approach to multiple regression analysis made by Sasaki (Ref 30). The approach determines the posterior distribution of B and σ from knowledge of the posterior distribution of B and σ obtained from a previous experiment. The distribution of the previous experiment is considered the prior probability distribution to be used in another experiment to derive a posterior distribution of B and σ.

Prior Information and Distribution. If the first experiment consisted of N_1 observations on all Y and X, denoted Y_1 and X_1 for the dependent and independent variables of experiment one. The following is the posterior distribution of B and σ.

$$P(B,\sigma_1|Y_1,X_1) = \alpha K(2\pi)^{-N_1/2}\sigma_1^{-(N_1+1)}$$

$$X \exp\left\{-\frac{1}{2\sigma_1^{2}}[\sigma_1^{2}\gamma_1 + (B-b')'X'_1X_1(B-b')]\right\} \qquad (53)$$

where b' is the least-squares estimate, and γ_1 is the number of degrees of freedom from this first sample. A detailed derivation of the above formula is presented in Sasaki (Ref 30:464-467).

Current Information and Distribution. Now a new random sample of size N_2 is taken on all Y and X. The second set of observations are denoted Y_2 and X_2. The likelihood of getting these observations is

$$L(Y_2, X_2 | B_1 \sigma_2) = (2\pi)^{-N_2/2} \sigma_2^{-N_2} \exp\left\{ -\frac{1}{2\sigma_2^2} [(Y_2 - X_2 B)' (Y_2 - X_2 B)] \right\}$$

$$= (2\pi)^{-N_2/2} (\sigma_2^{-N_2}) \times \exp\left\{ -\frac{1}{2\sigma_2^2} [(Y_2 - X_2 b^{"})' (Y_2 - X_2 b^{"}) + B - b^{"})' X_2' X_2 (B - b^{"})] \right\}$$

$$= (2\pi)^{-N_2/2} (\sigma^{2-N_2})$$

$$\times \exp\left\{ -\frac{1}{2\sigma_2^2} [\hat{\sigma}_2^2 \gamma 2 + (B - b^{"})' X'_2 X_2 (B - b^{"})] \right\} \quad (54)$$

where

$\sigma_2^2 = (Y_2 - X_2 b^{"})' (Y_2 - X_2 b^{"})/(N_2 - k)$, and $b^{"}$ denotes the least-square estimate of B for the new sample (Ref 30:468).

Posterior Distribution and Estimate of B. The posterior distribution of B and σ is proportional to the product of the prior distribution and the likelihood function. The posterior distribution of these parameters can be derived from equations (53) and (54). The following is an expression for this product (Ref 30:468).

$$P(B,\sigma|Y_2,X_2,Y_1,X_1) = \text{Constant}\ \sigma^{-(N_1 + N_2 + 1)}$$

$$\exp\left\{-\frac{1}{2\sigma^2}\left[\hat{\sigma}_2^{\,2}\gamma_2 + \hat{\sigma}_1^{\,2}\gamma_1 + (B-b)'X'X(B-b)\right]\right\} \tag{55}$$

The assumption made is that $\sigma_1 = \sigma_2 = \sigma$, therefore

$$X'X = X_1'X_1 + X_2'X_2 \tag{56}$$

and b the posterior estimate of B is

$$b = (X'X)^{-1}(X_1'X_1b' + X_2'Y_2b'') = (X'X)^{-1}(X_1'Y_1 + X_2'Y_2) \tag{57}$$

<u>Posterior Variance</u>. By integrating equation (55) with respect to σ, the marginal posterior distribution can be derived along with an estimate of σ. The result is

$$P(B|Y_2,X_2,Y_1,X_1) = \text{Constant}\left\{1 + (B-b)'X'X(B-b)/\hat{\sigma}^2\gamma\right\}^{-(\gamma+k)/2} \tag{58}$$

where

$$\hat{\sigma}^2 = \frac{\hat{\sigma}_1^{\,2}\gamma_1 + \hat{\sigma}_2^{\,2}\gamma_2}{\gamma} \tag{59}$$

and

$$\gamma = N_1 + N_2 - K \tag{60}$$

which is the form for the multivariate t distribution with γ degrees of freedom (Ref 30:468-472).

Translation of the Methodologies to Updating Airframe CERs

As discussed in Chapter II, the RANDOM CERs are based on a mixed linear model having two error terms. Although Chapter II provides some discussion of the application of a mixed linear model for development of an airframe CER, more detail concerning the use of GLS techniques in the development of the RANDOM CER is required to provide the reader with an understanding of the methodology used to obtain updated CERs.

RANDOM CER. The general model for the RANDOM CER was a mixed linear model or components of variance model,

$$Y_{ij} = X_{ij}B + u_j + \varepsilon_{ij} \tag{61}$$

where, Y is a vector of cost, X is the matrix of explanatory variables of weight, speed, and quantity, u_j is the error term associated with a particular aircraft program, and ε_{ij} is the error across all lots of all aircraft aircraft programs. The error associated with different aircraft airframe programs may result from several factors such as totally new aircraft designs, new materials used, and new and complex technical design features, etc. Aircraft airframe differences definitely affect cost, but the explanatory variables used in the CER may not capture totally the variance resulting from these differences. The general economic conditions present during production of each lot of airframes is the most probable cause of variability reflected by the overall error term, ε_{ij}. This error may be due to small changes in labor or material caused by strikes and inflation or slight changes in the designs that occur in aircraft airframe programs between lots (Ref 19:13).

Marcotte assumed that the distribution of the error terms were normal, independent, and had variances of σ_u^2 and σ_ε^2, respectively. As a result of the assumptions, the overall error could be presented as

$$g_{ij} = u_j + \varepsilon_{ij} \qquad (62)$$

The overall g_{ij} is also distributed normally with a mean α and a variance equal to the sum of the variances, $\sigma_u^2 + \hat{\sigma}_\varepsilon^2$. Further development resulted in the general model,

$$Y = XB + G \qquad (63)$$

where G is distributed normally with a mean and a variance matrix of α and V, respectively. Since the G disturbance terms are correlated, the V matrix of variance will have non-zero off diagonal terms (Ref 19:13,14).

Marcotte applied the fitting constants technique to the mixed linear model to obtain estimates for σ_u^2 and σ_ε^2. With values of $\hat{\sigma}_u^2$ and $\hat{\sigma}_\varepsilon^2$, a V matrix can be constructed (Ref 19:11-18).

The estimated $\hat{V}$ matrix is a positive definite symmetric matrix. Using triangularization of a matrix, a subroutine of the *OMNITAB II computer program*, a nonsingular matrix (T) was determined such that

$$T'T = V^{-1} \tag{64}$$

where V is Q and T represents the P^{-1} of equation (33). Using the GLS methodology discussed earlier, the dependent variable vector of costs and the matrix of independent variables (weight, speed, and quantity) from the historical aircraft airframe data was multipled by the T matrix. Ordinary regression was performed on the resulting products formed to obtain estimates of regression coefficients and variances for the development of the RANDOM CERs. These estimates and the historical data base to derive the estimates are the prior information for application of a Bayesian approach to multiple regression, while the available or observed data on a new program is the current information.

Differences in Notation. In the above paragraph, a difference in notation was pointed out to the reader. Clarification of the differences in notation between Johnston, Marcotte and this research effort are important, in order to follow the translation of the methodology to updating the RANDOM CER.

The $\hat{V}$ matrix estimated by Marcotte is not equalivalent to the V matrix presented in Johnston, equations

(50), (51), and (52). The Johnston V matrix is assumed to be a known symmetric, positive definite matrix and is equivalent to $\sigma^2 Q$, if σ^2 is known, or $\sigma^2 \hat{V}^{-1}$ using the notation employed by Marcotte. To obtain an estimate of the V matrix, σ^2 or average variance across all observations must be estimated using equation (40). In this effort the notation for the estimate of the average variance (σ^2) or adjustment to the $\hat{V}^{-1}$ matrix for prediction will be δ^2. Furthermore, $\delta^2 W$ is an estimate of the W vector in equation (45). Refer to Theil, Principles of Econometrics (Ref 35) for another presentation of GLS.

Posterior $\hat{B}$ Estimate. An ordinary regression analysis was conducted on the summation of the new experiment data and the historical data of an example presented by Sasaki (Ref 30) to verify that the coefficients of explanatory variables, $\hat{B}$ estimates, calculated were the same as the posterior $\hat{B}$ estimates using equation (57). Since the estimates were equivalent and the method simplifies the methodology presented by Sasaki for obtaining posterior estimates of $\hat{B}$, this simplier method was used.

The logarithm of the dependent variables of weight, speed, and quantity of the 33 lots of historical data (X_1) and the four lots of observed data (X_2) were added together, represented by the vector

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \tag{65}$$

The logarithm of the independent variable of average unit cost for marginal equations or cumulative lot cost for cumulative equations for both sets of data is represented by the vector

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \tag{66}$$

The non singular matrix

$$T = \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix} \tag{67}$$

was derived by Marcotte using partitioning of the T matrix.

Applying GLS the posterior estimates of B were derived using equation (36),

$$\hat{B} = (X'T'TX)^{-1}X'T'Y \tag{68}$$

The variance ($\hat{\delta}^2$) resulting from the regression performed on the combined data set is the estimate of the avarage variance, and an adjustment to the $\hat{V}^{-1}$ matrix for prediction.

Posterior Variance, $\hat{\sigma}^2$. Using GLS, ordinary regression was performed on the historical data represented by the following equation,

$$T_1Y_1 = T_1X_1B + T_1G_1 \tag{69}$$

An estimate of variance ($\hat{\sigma}_1{}^2$) was obtained from this regression.

Applying GLS and performing an ordinary regression on the observed data for the new program represented by the following equation,

$$T_2Y_2 = T_2X_2B + T_2G_2 \tag{70}$$

results in an estimate of variance ($\hat{\sigma}_2{}^2$) for this observed data.

Now that estimates for $\hat{\sigma}_1{}^2$ and $\hat{\sigma}_2{}^2$ have been calculated, the posterior variance ($\hat{\sigma}^2$) is calculated using equation (59).

Prediction. Using the Bayesian updated RANDOM CER to predict the cost of the next lot of airframes, presents a major problem. The problem is how to take into consideration the correlation that exists in an airframe program between the observed lots of airframes and the lot of airframes for which a cost estimate is needed to make a management decision. The RANDOM CER was developed based on the assumption that correlation exists between the produced lots of a given airframe program. Therefore, GLS treatment of prediction is applicable.

The form of the Bayesian updated RANDOM CER is

$$C = e^{\alpha} S^{\hat{B}_1} W^{\hat{B}_2} Q^{\hat{B}_3} \qquad (71)$$

where,

C = Lot Cost.

α = Coefficient of the constant term.

$\hat{B}_1$, $\hat{B}_2$, $\hat{B}_3$ = Posterior regression coefficients.

S = Maximum speed at best altitude.

W = Unit airframe weight.

Q = Airframe quantity.

To simplify the presentation of predicting airframe cost using the Bayesian updated CER, the following linear equation represents the prediction equation.

$$Y_0 = X_0\hat{B} + \varepsilon_0 \qquad (72)$$

where, Y_0 is the logarithm of the cost to be predicted, X_0 is a row vector composed of the logarithm of the explanatory variables (S, W, and Q) for the airframe for which the cost is to be predicted, $\hat{B}$ is a column vector of the posterior estimates of the regression coefficients, and ε_0 is the unknown, but true value of the prediction disturbance. The Bayesian updated RANDOM CER has been converted to the form presented in equation (42). The same assumptions hold.

The covariances between the prediction disturbance and the sample disturbances that compare to the column vector W, equation (31), are assumed to be constant and equivalent to the covariance between to different lots of the same airframe, $\hat{\sigma}_u^2$. The values used for $\hat{\sigma}_u^2$ were estimated by Marcotte. The number of rows which make up the n X 1, W vector, is equivalent to the number of lot observations available on the new program used to predict the cost of the next lot to be produced. For example, if the cost of the fourth lot of airframes is to be predicted W would consist of three rows, one for each of the first three lot observations.

Substitution of $T'T$ for V^{-1} in equation (51) results in

$$\hat{C} = T'T(I - X(X'T'TX)^{-1}X'T'T)W + T'TX(X'T'TX)^{-1}X_0' \quad (73)$$

The estimate of the predictor, $\hat{p}$, becomes

$$\hat{p} = \hat{C}'Y \quad (74)$$

$$= W' - (I - T'TX(X'T'TX)^{-1}X')T'TY + X_0(X'T'TX)^{-1}X'T'TY \quad (75)$$

$$= W'T'TY - W'T'TX\hat{B} + X_0\hat{B} \quad (76)$$

$$= W'T'(TY - TX\hat{B}) + X_0\hat{B} \quad (77)$$

The $TY - TX\hat{B}$ portion of equation (77) is made up of the residuals associated with the available lot information on the new program multipled by the T matrix. The value for the products of the vector of residuals times T were obtained from the regression analysis performed on the combined data set of historical data and the available observations on the new aircraft program. The product $W'T'(TY - TXB)$ is equivalent to $W'V^{-1}\varepsilon$ and is an estimate of the unknown prediction disturbance ε_0. The $X_0\hat{B}$ portion is a vector of the values of the variables of weight, speed, and quantity for the lot of airframes to be predicted times the vector of Bayesian updated regression coefficients. Equation (52), when converted from the log-linear form becomes

$$C = e^{\alpha} S^{\hat{B}_1} W^{\hat{B}_2} Q^{\hat{B}_3} e^{K} \qquad (78)$$

where $K = W'V^{-1}\varepsilon$, the estimate of the unknown prediction disturbance in the log-linear form of the equation and is the coefficient of an additional constant term or adjustment factor. This additional constant term takes into consideration the correlation among the lot observations of the new program.

Now that the prediction equation has been developed, the standard deviation of the prediction (σp) must be estimated to calculate a 95 percent prediction interval.

From equation (50),

$$\sigma_p^2 - \sigma_0^2 = C'\delta^2 VC - 2C'\delta^2 W \qquad (79)$$

$$\text{let } a = \delta^2 (C'VC - 2C'W) \tag{80}$$

$$= - W'T'TW\delta^2$$

$$- 2X\delta^2 (X'T'TX)^{-1}X'T'TW$$

$$+ W'T' [TX\delta^2 (X'T'TX)^{-1}X'T'](T')^{-1}W$$

$$+ X_0\delta^2 (X'T'TX)^{-1}X_0' \tag{81}$$

Now, assume that variance is constant, $\hat{\sigma}_0^2 = \hat{\sigma}^2 = \hat{\sigma}^2$ then

$$\hat{\sigma}_p^2 = a + \hat{\sigma}^2 \tag{82}$$

and

$$\hat{\sigma}_p = \sqrt{a + \hat{\sigma}^2} \tag{83}$$

The prediction interval is the range between estimated limits in which a value of a single point estimate lies with some probability. The prediction interval also can be described as the range of the conditional distribution within which most individual values fall (Ref 30:420).

Assuming the following equation

$$Y_i = a + BX_i = \varepsilon_i \quad (84$$

and a large sample size, the interval estimate for 95% confidence is

$$P\ [(a + BX_0) \pm Z_{\frac{.05}{2}} \frac{\sigma}{\sqrt{N}}] = 0.95 \tag{85}$$

For a small sample size the t distribution is used. The interval estimate for 95% confidence is then

$$P[(a + BX_0) \pm t_{0.25} \sqrt{VAR\ (\hat{Y}_0)}] = 0.95 \tag{86}$$

Because the data base used to derive the Bayesian equations was a small sample, the t value for infinite degrees of freedom was used to calculate the upper and lower limits of a 95% prediction interval. An estimate of VAR $\hat{Y}_0$ has been calculated and is equivalent to $\hat{\sigma}_p^{\,2}$. Due to the logarithmic form of the CER, the expression for the prediction interval is

$$P\ [\ln \hat{Y} \pm 1.96\ \hat{\sigma}_p] = 0.95 \tag{87}$$

The upper and lower limits are found by calculating $e^{(\ln Y \pm \ln 1.96\ \hat{\sigma}_p)}$.

Summary

This chapter discussed the methodology employed to obtain updated airframe CERs. To familiarize the reader with Bayesian statistics, the chapter began with a discussion of Bayesian philosophy. An example of Bayes Theory applied to a discrete case was presented and discussed. In addition, two concepts were discussed which are significant to the development of the Bayesian approach used in this thesis. The first concept discussed was Generalized Least-Squares (GLS). The use of GLS in the methodology was necessary due to an underlying assumption of the RANDOM

CERs of the existence of correlation between two error terms. The second concept discussed was a Bayesian approach to multiverate regression analysis. The final section translates the use of these concepts for updating CERs (Bayesian CERs) to predict costs of the next lot/ unit of airframes. Also included in the final section is a brief discussion for calculating 95% confidence intervals for the Bayesian predictions. The following chapter presents a comparison of the Bayesian CERs and predictions with those developed using other methodologies. The 95% confidence interval for the Bayesian predictions are also presented.

V. Equations and Comparisons

This chapter presents the marginal and cumulative Bayesian equations developed for predicting the cost of the next airframe unit or lot, and compares the Bayesian equation prediction with predictions from equations derived using other methodologies. The other equations used for the comparisons are the Levenson equation (Timson and Tihansky equations developed by Marcotte), RANDOM equations, ordinary regression equations (lot observations assumed independent) and Large equations. The first section discusses the problems of using the F-14 program as the sole test case. The next section of the chapter compares the RANDOM equations and ordinary regression equations with the Bayesian equations. Comparisons are made for both marginal and cumulative cost equations for each of the following categories: total engineering, total tooling, recurring labor, and recurring material. The last section presents the Bayesian predictions and comparisons with predictions made using the other equations mentioned previously.

Test Case Problems

The Bayesian equations using the F-14 program as the test case of the Bayesian methodology were used to predict the recurring material fifth lot cost. Performance of ordinary regression on the observed lot information raised questions concerning the selection of the F-14 program as the test case of the Bayesian methodology, because the learning-curve slope (quantity coefficient) of the material cost category differs significantly from the average learning-curve slope indicated by the regression on the historical data, RANDOM equation. The reasons for this significant difference may be because the airspeed was estimated from published information or the assumption of the acceptance schedule. The estimate of airspeed did not pose a problem except for the engineering category since the airspeed variable was not significant for the other categories. For an explanatory variable to be significant the t-statistic must be greater than ± 2.0. The t-statistics of the airspeed variable in the cumulative total engineering cost equations were not much greater than ± 2.0. The assumed acceptance schedule is a possible source of error. Although the acceptance schedule assumption only affects the recurring material cost category, the A-6 airframe program was selected as a second test case.

Equations

The following tables compare the coefficients of the explanatory variables, the estimate of variance ($\hat{\sigma}^2$), and the standard deviation of prediction for the i^{th} unit ($\hat{\sigma}_{pi}$) for the equations derived using the Bayesian, RANDOM, and ordinary regression methodologies.

The RANDOM equations were derived by Marcotte using 33 lot observations of fighter aircraft. The lot observations of the test case airframe programs were not included in the 33 lot observations. The Bayesian equations to predict the cost of the fifth lot were derived using the 33 lots of fighter aircraft data and four lot observations of the A-6 or F-14 airframe programs. The ordinary regression equations to predict the cost of the fifth lot were derived using only four lots of A-6 or F-14 information.

Because the recurring labor category is defined better than the other cost categories, the recurring labor cost data is more consistent and comparable. The labor category is the only cost category which legitimately reflects a "true" learning-curve. For the above reasons, the recurring labor cost category was selected for further analysis using the Bayesian methodology. In addition to the recurring labor equations developed using four observations to predict the labor cost of the next lot, Bayesian equations were derived to predict the labor cost of the next lot using one, two and three lots of information. An ordinary regression equation was developed using three lots of information and a straight live equation using two lots of information. These equations were derived for comparison purposes.

The following are the definitions of the explanatory variables for the coefficients presented in the tables:

K_i - Coefficient of the constant term (e^{K_i}).

S - Maximum speed at best altitude (knots).

W - Unit weight (pounds).

Q - Quantity of airframes (cumulative total or true lot midpoint).

K_2 - Coefficient of the Bayesian adjustment constant (e^{K_2}).

The numbers enclosed in the parentheses below the value of the coefficient are the t-statistic for the coefficient. The relative magnitude of the t-statistic greater than +2.0 indicates the relative significance of the particular explanatory variable in explaining the cost.

Cumulative Equations. Tables VIII-XI present the cumulative cost equations derived for total engineering, total tooling, recurring labor, and recurring material, respectively. The statistics presented are the variance ($\hat{\sigma}^2$), standard deviation of prediction for the fifth lot of airframes ($\hat{\sigma}_{p5}$), and in parentheses the t-statistic.

The ordinary regression equations consistenly reflect a smaller value of $\hat{\sigma}_{p5}$ than the Bayesian equations. The value of $\hat{\sigma}_{p5}$ should not be used for determining which methodology provides the better prediction equation since the regression equation was derived using four lots of information. The lower value of $\hat{\sigma}_{p5}$ may only reflect a lack of variability in the data.

Another observation is $\hat{\sigma}^2$ of the Bayesian equations are less than the value for the RANDOM equation in all four tables. Also, the coefficient of the quantity variable is adjusted in the Bayesian equation from the value in the RANDOM equation towards a value closer to the coefficient of quantity in the regression equation. This was observed in all four tables. For example, the RANDOM CER for cumulative recurring labor hours in Table X has a value of .6104 for the coefficient of quantity. The regression equation results in a coefficient of quantity of .7481. These values indicate that the A-6 program has a different learning-curve slope than the average learning-curve slope represented by the coefficient of quantity of the RANDOM equation derived from the historical data base. Using the Bayesian methodology the value of the coefficient of quantity changes from the RANDOM CER value, .6104, to a value of .6353 in the Bayesian equation. Both the decreasing variance and changing coefficient of quantity illustrate the weighting effect of the Bayesian methodology on prior information as new information is considered.

For the cumulative RANDOM and Bayesian engineering CERs in Table VIII, the t-statistic indicates that all three variables of speed, weight, and quantity are significant in determining the engineering cost. An explanation for this occurring is that the physical performance characteristics represented by the three variables directly relate to the technology employed. The technology employed, state-of-the-art or advanced, is directly related to the total engineering cost.

The t-statistic of the explanatory variables of the cumulative RANDOM and Bayesian tooling CERs, Table IX, indicate quantity to be a significant variable for determining the cost of this category. This result should be expected for this category since tooling costs are related to the number of machine set ups required and the length of the production run for amortization of the tooling costs.

The t-statistic of the explanatory variables of the cumulative RANDOM and Bayesian CERs, Table X, indicates weight and quantity are significant. The significance of quantity in the labor category equation is expected since the characteristic of learning associated with labor has been found to be directly related to the cost of labor. As more units are produced, labor cost decreases. The significance of weight may be attributed to a relationship that heavier airframes are larger, therefore requiring more manufacturing labor resulting in increased labor costs.

The t-statistic of the explanatory variables of the cumulative RANDOM and Bayesian material CERs, Table XI, indicates that quantity and weight are significant variables for predicting cost. The significance of quantity should be expected for this category. Material costs are directly related to quantity discounts, improved manufacturing techniques, and learning attributed to labor working with the material. Using the magnitude of the t-statistic, weight was considerably less significant than quantity, but did

Table VIII
CER Coefficients Cumulative Engineering
for Fighter Aircraft

CER Methodology	Aircraft Type	Estimated Coefficients for Explanatory Variables K_1	S	W	Q	K_2	$\hat{\sigma}^2$	$\hat{\sigma}_{p5}$
Random	All	-.7931 (-.55)	.7859 (2.56)	1.041 (7.54)	.2476 (12.35)	-	1.0238	
Bayesian	A-6	-.0351 (-.03)	.5521 (2.48)	1.1444 (9.90)	.2234 (11.65)	.1008	.9007	.9334
	F-14	-1.1278 (-.79)	.7587 (2.48)	1.1075 (8.33)	.2303 (13.25)	.2132	.9046	.9372
Ordinary Regression (4 Observations)	A-6	15.0760 (775.97)	-	-	.1173 (20.61)	-	.0001	.0140
	F-14	15.8832 (462.61)	-	-	.1922 (18.24)	-	.0005	.0308

Table IX
CER Coefficients Cumulative Tooling for
Fighter Aircraft

CER Methodology	Aircraft Type	Estimated Coefficients for Explanatory Variables K_1	S	W	Q	K_2	$\hat{\sigma}^2$	$\hat{\sigma}_{p5}$
Random	**All**	9.6203 (2.40)	-.1040 (-.13)	.5669 (1.70)	.3135 (11.60)	-	.8631	
Bayesian	**A-6**	7.6622 (2.34)	.3991 (.73)	.4044 (1.52)	.3136 (13.35)	-.2974	.7779	.7520
	F-14	9.7019 (2.51)	-.1783 (-.23)	.6227 (1.97)	.2928 (12.67)	.0029	.7597	.7355
Ordinary Regression (4 Observations)	**A-6**	13.8915 (112.44)	-	-	.2955 (8.17)	-	.0039	.0894
	F-14	14.9990 (612.73)	-	-	.2221 (29.56)	-	.0002	.0219

Table X
CER Coefficients Cumulative Labor
for Fighter Aircraft

CER Methodology	Aircraft Type	Estimated Coefficients for Explanatory Variables K_1	S	W	Q	K_2	$\hat{\sigma}^2$	$\hat{\sigma}_{p5}$
Random	All	4.1595 (1.69)	-.1704 (-.34)	1.1093 (5.27)	.6104 (33.70)	-	.9418	
Bayesian	A-6	3.2542 (1.43)	.1215 (.32)	.9787 (5.17)	.6353 (35.55)	-.1223	.8347	.8558
	F-14	4.3768 (1.86)	-.1360 (-.28)	1.0533 (5.33)	.6226 (40.59)	-.1567	.8296	.8656
Ordinary Regression (4 Observations)	A-6	13.0694 (272.06)	-	-	.7481 (53.18)	-	.0006	.0347
	F-14	13.8663 (678.86)	-	-	.6593 (105.17)	-	.0002	.0183

Table XI
CER Coefficients Cumulative Material
for Fighter Aircraft

CER Methodology	Aircraft Type	Estimated Coefficients for Explanatory Variables					$\hat{\sigma}^2$	$\hat{\sigma}_{p5}$
		K_1	S	W	Q	K_2		
Random	All	3.5994 (1.19)	.6885 (1.11)	.6508 (2.41)	.7976 (29.85)	-	.9861	
Bayesian	A-6	3.2067 (1.27)	.8267 (1.94)	.5814 (2.69)	.8174 (34.32)	-.0499	.8701	.8645
	F-14	3.1412 (.69)	.4824 (.51)	.8988 (2.26)	.7079 (19.95)	.4205	.8770	.7655
Ordinary Regression (4 Observations)	A-6	13.7524 (279.27)	-	-	.9105 (63.14)	-	.0006	.0826
	F-14	17.050 (251.43)	-	-	.4171 (20.04)	-	.0018	.0608

meet the ±2.0 criteria. The writer can not provide an explanation for the significance of the weight variable.

A possible and very likely explanation of why weight and speed are considerably less significant than quantity for these equations is the homogeneous nature of the data base. The data base consists of a specific type of airframe, fighters, having a narrow range of variability for the physical performance variables of weight and speed. Therefore, quantity produced becomes the more significant variable for determining cost.

Marginal Equations. Tables XII-XV present the marginal equations derived to estimate total engineering, total tooling, recurring labor and recurring material costs. The tables present the coefficient of the explanatory variables of the equations developed using the RANDOM, Bayesian, and ordinary regression methodologies. The Bayesian and regression equations were developed using information about four lots of the A-6 or F-14 airframe program to predict the cost of the fifth lot. The t-statistic is presented in parentheses below the coefficient.

The ordinary regression equations have a lower $\hat{\sigma}_{p5}$ than the Bayesian equations for all cost categories. As stated before the value of $\hat{\sigma}_{p5}$ should not be used as the criteria for determining the better prediction methodology. The regression equations were derived using only four lots of information, therefore, the lower $\hat{\sigma}_{p5}$ may reflect the lack of any significant variability in the data.

Table XII
CER Coefficients (Marginal) Engineering
for Fighter Aircra

CER Methodology	Aircraft Type	Estimated Coefficients for Explanatory Variables K_1	S	W	Q	K_2	$\hat{\sigma}^2$	$\hat{\sigma}_{p5}$
Random	All	-2.8788 (-.97)	1.0905 (1.69)	1.0024 (3.29)	-.9732 (-10.29)	-	1.0027	
Bayesian	A-6	-3.7521 (-1.49)	1.3502 (3.14)	.9146 (3.96)	-.9959 (-11.71)	-.0269	.9232	.9349
	F-14	-2.6816 (-.96)	1.1049 (1.81)	.9725 (3.53)	-.9795 (-12.80)	-.0296	.8845	.9467
Ordinary Regression	A-6	14.3859 (21.62)	-	-	-1.2732 (-5.79)	-	.2392	.6774
	F-14	15.4132 (91.82)	-	-	-1.1578 (-19.64)	-	.0189	.1969

Table XIII
CER Coefficients (Marginal) Tooling
for Fighter Aircraft

CER Methodology	Aircraft Type	Estimated Coefficients for Explanatory Variables K_1	S	W	Q	K_2	$\hat{\sigma}^2$	$\hat{\sigma}_{p5}$
Random	All	11.0981 (2.34)	-1.0051 (-.98)	1.0418 (2.21)	-.8974 (-11.71)	-	.8701	
Bayesian	A-6	6.6607 (1.66)	.3259 (.47)	.5328 (1.45)	-.8912 (-12.57)	-.5183	.7773	.7596
	F-14	11.7680 (2.61)	-.9716 (-.98)	.9456 (2.17)	-.9061 (-13.52)	-.3621	.7705	.7624
Ordinary Regression (4 Observations)	A-6	13.6106 (44.12)	-	-	-1.0015 (-9.34)	-	.0499	.3103
	F-14	14.4978 (78.02)	-	-	-1.0730 (-16.46)	-	.0230	.2176

Table XIV
CER Coefficients (Marginal) Labor
for Fighter Aircraft

CER Methodology	Aircraft Type	Estimated Coefficients for Explanatory Variables K_1	S	W	Q	K_2	$\hat{\sigma}^2$	$\hat{\sigma}_{p5}$
Random	All	4.3113 (2.20)	-.6165 (-1.45)	1.3781 (7.03)	-.3958 (-10.39)	-	.8869	
Bayesian	A-6	2.6416 (1.59)	-.1025 (-.36)	1.1683 (7.72)	-.3740 (-10.71)	-.1846	.7810	.8677
	F-14	4.6966 (2.52)	-.5916 (-1.45)	1.3105 (7.24)	-.3822 (-11.66)	-.1857	.7868	.8714
Ordinary Regression (4 Observations)	A-6	12.9292 (238.84)	-	-	-.2975 (-16.77)	-	.0014	.0524
	F-14	13.6011 (135.60)	-	-	-.3803 (-10.88)	-	.0063	.1145

Table XV
CER Coefficients (Marginal) Material
for Fighter Aircraft

CER Methodology	Aircraft Type	Estimated Coefficients for Explanatory Variables						
		K_1	S	W	Q	K_2	$\hat{\sigma}^2$	$\hat{\sigma}_{p5}$
Random	All	3.8216 (1.68)	.1795 (.37)	.9771 (4.38)	-.2029 (-6.51)	-	.9426	
Bayesian	A-6	2.5528 (1.34)	.5727 (1.75)	.8144 (4.72)	-.1836 (-6.47)	-.1636	.8350	.8827
	F-14	3.6894 (1.15)	.1187 (.17)	1.068 (3.48)	-.2741 (-6.83)	.0576	.9997	.9394
Ordinary Regression (4 Observations)	A-6	13.6642 (162.16)	-	-	-.0880 (-3.19)	-	.0033	.0807
	F-14	16.4748 (45.49)	-	-	-.6568 (-5.18)	-	.0852	.4190

The observations concerning the weighting effect of the Bayesian approach and the significance of the t-statistic are the same as discussed for the cumulative equations. The result which led to the selection of the A-6 program as a second test case is illustrated by the disparity between the coefficient of quantity for the RANDOM equation and the regression equation for the F-14 airframe program found in Table XV. The value of the RANDOM equation is -.2029 and reflects an average fighter airframe program learning-curve. The regression equation for the F-14 airframe program is -.6568 reflecting a higher percentage learning. This significant difference resulted in the decision to select a second airframe program as a test case.

Analysis of Recurring Labor. The recurring labor cost category was selected for a more detailed analysis for determination of the predictive capability of the Bayesian methodology. The reasons for selecting this category were discussed at the beginning of this section.

Bayesian equations were developed using one, two, or three lots of information to predict the cost of the next lot of airframes. The Bayesian equations derived using the A-6 and F-14 observations are compared with a regression equation developed using three lots of information and a straight line equation developed using two lots of information.

The coefficients of the explanatory variables for the cumulative equations developed to predict the recurring labor cost of the second, third, and fourth lot of A-6 and F-14 are presented in Tables XVI and XVII, respectively. In addition to the coefficients, the t-statistic in parenthesis, the variance ($\hat{\sigma}^2$), standard deviation of prediction for the next lot ($\hat{\sigma}_{pi}$), and the standard deviation of prediction of the succeeding lots through the fifth lot are presented.

The variance for the Bayesian equations for both the A-6 and F-14 programs decreases as additional lot information is considered. For example, the variance for the cumulative Bayesian equation for the A-6 program using the first lot observation was .9104. When three lots of information were used, the variance decreases to .8562. For four lots of information, Table X, the variance decreases to .8347. This result can be expected due to the weighting of variance using the Bayesian methodology. As more observations of a new program are used, the variance of the new program will carry an increasing weight due to the increase in the degrees of freedom. In addition, the lot observations of the new program consists of one type of airframe; less variability between the lot observations exists than is present in the historical data.

Table XVI
Equations and Standard Deviation of Prediction for Cumulative Labor Hours for Less than Four Observations on the A-6

Number of Observations	CER Methodology	Estimated Coefficients for Explanatory Variables					Estimates of Std Dev of Prediction*				
		K_1	S	W	Q	K_2	$\hat{\sigma}^2$	$\hat{\sigma}_{p2}$	$\hat{\sigma}_{p3}$	$\hat{\sigma}_{p4}$	$\hat{\sigma}_{p5}$
3 Lots	Bayesian	3.0747 (1.38)	.1492 (.40)	.9817 (5.23)	.6264 (34.51)	-.1541	.8562	-	-	.8716	.8719
	Regression	13.0260 (248.39)	-	-	.7647 (44.11)	-	.0004	-	-	.0345	.0406
2 Lots	Bayesian	2.8508 (1.32)	.1912 (.53)	.9781 (5.49)	.6187 (34.24)	-.1919	.8810	-	.8897	.8900	.8903
	Straight Line	12.9627	-	-	.7915	-	@	-	-	-	-
1 Lot	Bayesian	2.5926 (1.23)	.2504 (.70)	.9638 (5.52)	.6148 (34.36)	-.2327	.9104	.9092	.9095	.9099	.9103
	One Point (Actual Observation)	-	-	-	-	-	@	-	-	-	-

*Subscript represents prediction for follow-on lots from early actual observations.

@Variance cannot be estimated.

Table XVII
Equations and Standard Deviation of Prediction for Cumulative Labor Hours for Less than Four Observations on the F-14

Number of Observations	CER Methodology	Estimated Coefficients for Explanatory Variables					Estimates of Std Dev of Prediction*				
		K_1	S	W	Q	K_2	$\hat{\sigma}^2$	$\hat{\sigma}_{p2}$	$\hat{\sigma}_{p3}$	$\hat{\sigma}_{p4}$	$\hat{\sigma}_{p5}$
3 Lots	Bayesian	4.4500 (1.89)	-.1560 (-.33)	1.0628 (5.38)	.6165 (37.67)	-.1872	.8537	-	-	.8798	.8801
	Regression	13.8899 (1314.42)	-	-	.6492 (168.68)	-	.0000	-	-	.0091	.0105
2 Lots	Bayesian	4.4945 (1.89)	-.1656 (-.34)	1.0663 (5.34)	.6136 (35.56)	-.2076	.8810	-	.8946	.8951	.8954
	Straight Line	13.9097	-	-	.6396	-	@	-	-	-	-
1 Lot	Bayesian	4.4999 (1.86)	-.1660 (-.34)	1.0660 (5.25)	.6135 (34.81)	-.2106	.9104	.9103	.9107	.9113	.9117
	One Point (Actual Observation)	-	-	-	-	-	@	-	-	-	-

*Subscripts represent prediction for follow-on lots from early actual observations.

@Variance cannot be estimated.

Tables XVIII and XIX present the coefficients of the marginal recurring labor equations developed using one, two or three lots of A-6 or F-14 airframe data to update the RANDOM CER.

The same observations discussed for the cumulative equations in Tables XVI and XVII are found in the marginal equations and statistics.

Comparison of Predictive Capability

The previous section presented the comparison of Bayesian equations developed to predict future lot cost. This section will present the predicitons made using the Bayesian equations and compare these predictions with predictions made using other airframe CERs.

Measurement Used. The measurement used to compare the predictions is the percent deviation of the prediction from the actual cost. The percentage is calculated by subtracting the prediction from the actual cost and then dividing the result by the actual cost. The reason for using this measurement for comparison of predictive capability is due to the difficulty of deriving the statistics for the Bayesian equation which are normally used for comparison of predictability of regression equations. Using the percent deviation from actual value as the means of comparison has the advantages of being simple and straightforward.

Table XVIII
Equations and Standard Deviation of Prediction for (Marginal) Labor Hours for Less than Four Observations on the A-6

Number of Observations	CER Methodology	Estimated Coefficients for Explanatory Variables					$\hat{\sigma}^2$	Estimates of Std Dev of Prediction*			
		K_1	S	W	Q	K_2		$\hat{\sigma}_{p2}$	$\hat{\sigma}_{p3}$	$\hat{\sigma}_{p4}$	$\hat{\sigma}_{p5}$
3 Lots	Bayesian	2.6144 (1.54)	-.0949 (-.32)	1.1660 (7.53)	-.3749 (-10.34)	-.1875	.8046	-	-	.8821	.8835
	Regression	12.8984 (195.83)	-	-	-.2818 (-11.15)	-	.0015	-	-	.1327	.1375
2 Lots	Bayesian	2.6073 (1.49)	-.0928 (-.30)	1.1652 (7.29)	-.3750 (-10.07)	-.1882	.8296	-	.8984	.8999	.9015
	Straight Line	12.8465	-	-	-.2501	-	@	-	-	-	-
1 Lot	Bayesian	2.6468 (1.45)	-.1046 (-.31)	1.1696 (6.95)	-.3748 (-9.88)	-.1839	.9616	.9759	.9770	.9785	.9803
	One Point (Actual Observation)	-	-	-	-	-	@	-	-	-	-

*Subscript represents prediction for follow-on lots from early actual observations.

@Variance cannot be estimated.

Table XIX
Equations and Standard Deviation of Prediction for (Marginal) Labor Hours for Less than Four Observations on the F-14

Number of Observations	CER Methodology	Estimated Coefficients for Explanatory Variables K_1	S	W	Q	K_2	$\hat{\sigma}^2$	Estimates of Std Dev of Prediction* $\hat{\sigma}_{p2}$	$\hat{\sigma}_{p3}$	$\hat{\sigma}_{p4}$	$\hat{\sigma}_{p5}$
3 Lots	Bayesian	4.6994 (2.48)	-.5915 (-1.42)	1.3102 (7.12)	-.3824 (-11.25)	-.1872	.8062	-	-	.8829	.8842
	Regression	13.6804 (139.45)	-		-.4255 (-10.05)	-	.0041	-	-	.1111	.1311
2 Lots	Bayesian	4.6523 (2.42)	-.5939 (-1.41)	1.3167 (7.04)	-.3810 (-11.03)	-.1628	.8296	-	.8970	.8983	.8998
	Straight Line	13.3749	-	-	-.3120	-	@	-	-	-	-
1 Lot	Bayesian	4.5301 (2.34)	-.6013 (-1.42)	1.3367 (7.07)	-.3828 (-11.02)	-.1026	.8785	.9284	.9284	.9296	.9309
	One Point (Actual Observation)	-	-	-	-	-	@	-	-	-	-

*Subscript represents prediction for follow-on lots from early actual observations.
@Variance cannot be estimated.

Marginal versus Cumulative Equations. Bayesian equations were developed to estimate both the marginal and cumulative costs. Several reasons exists for developing the marginal or unit CER. First, the cumulative data may hide variation. The unit curve is much more responsive to a trend change, therefore, use of a marginal CER should improve the prediction of the next unit (Ref 6:69). Another reason involves the purpose of the RANDOM approach which was to acknowledge the fact that correlation may exist in the data between lots of the same type of airframe. Using cumulative data, another type of correlation takes place since successive lot costs include costs of preceding lot(s). This type of correlation is referred to as auto-correlation which bias the regression statistic, R^2. The biased statistic may indicate the cumulative CER fits the data better than it actually does. The marginal (unit) CER eliminates the presents of auto-correlation within the cost data (Ref 19:27). Due to these reasons and the nature of the problem, prediction of the next unit or lot of airframes based on all known information, the predictions using the marginal equations will be used to evaluate the predictive capability of the Bayesian approach.

Bayesian Predictions and Percent Deviation. Table XX presents the predictions made using the Bayesian equations and the upper and lower limits of the 95% prediction interval. The percentage column is the percentage deviation of the upper or lower limit from the prediction. This percentage value is used as a means of comparing the variance, and is calculated by subtracting the prediction from the estimate of the upper or lower limit. The resultant sum is divided by the prediction.

The upper and lower limits are consistent for both the marginal and cumulative cost predictions made for all categories except material. The percentage values in Table XX illustrate this result. For example, the marginal material category for the A-6 program has a 95% upper and lower limit prediction deviation of +81 percent and -45 percent, respectively. These values are consistent with the percentage deviation for the upper and lower limit percentages of 79 and -44 for the cumulative material cost prediction for the A-6 program. The 95% prediction interval for marginal material cost of the F-14 program results in an upper and lower limit percentage of 88 and -47 and for cumulative material cost 67 and -40 respectively. This cost category for the F-14 indicates some inconsistency between marginal and cumulative prediction. This again illustrates the problems with the F-14 material cost data.

Another observation made while analyzing the above phenomenon was that a significant difference exists between the value of the Bayesian coefficient of the quantity term for the F-14 cumulative material cost equation, .7079, and the value of the quantity coefficient of the RANDOM

Table XX
95% Prediction Interval
For Bayesian Equations

Program/ Lot #		Lower Limit	Percent	Prediction	Percent	Upper Limit
Cumulative Cost Data*						
Engineering						
A-6	Lot 5	3897	-47	7304	87	13688
F-14	Lot 5	11427	-47	21470	88	40340
Tooling						
A-6	Lot 5	2816	-40	4671	66	7748
F-14	Lot 5	6712	-39	11011	64	18063
Labor						
A-6	Lot 5	8457	-44	15042	78	26756
	Lot 4	5924	-44	10650	80	19146
	Lot 3	3740	-45	6806	82	12385
	Lot 2	2132	-46	3931	84	7248
F-14	Lot 5	13923	-44	24929	79	44636
	Lot 4	10155	-45	18358	81	33186
	Lot 3	5962	-45	10885	83	19874
	Lot 2	2898	-46	5348	85	9868
Material Cost						
A-6	Lot 5	37703	-44	67457	79	120693
F-14	Lot 5	196068	-40	328193	67	549353
Marginal Cost Data						
Engineering						
A-6	Lot 5	4500	-48	8737	94	16950
F-14	Lot 5	18210	-47	34439	89	65120
Tooling						
A-6	Lot 5	6370	-40	10626	67	17720
F-14	Lot 5	11750	-40	19628	67	32790
Labor						
A-6	Lot 5	53090	-44	95185	79	170670
	Lot 4	63590	-45	115126	81	208440
	Lot 3	87730	-45	160591	83	293960
	Lot 2	104700	-48	201911	93	389380
F-14	Lot 5	78710	-44	141484	80	254320
	Lot 4	98050	-45	177620	81	321750
	Lot 3	130730	-45	239063	83	437190
	Lot 2	217940	-46	407069	87	760320

*In Thousands of Dollars

Table XX (Continued)

Program/ Lot #		Lower Limit	Percent	Prediction	Percent	Upper Limit
Material						
A-6	Lot 5	271250	-45	491290	81	889840
F-14	Lot 5	762240	-47	1362786	88	2564330

*In Thousands of Dollars

NOTE: For each case the Acutal Value is within the prediction interval.

cumulative material equation and the Bayesian equation for the A-6 program, .7976 and .8174 respectively. A significant difference in the material learning-curves of the F-14 program and the A-6 program is illustrated in the cumulative regression equations for the material cost category.

The following reasons may explain the observations discussed in the above paragraphs. The first reason for the difference in the material learning-curve slopes is the technology advances made between the F-14 program and the A-6 and other airframe programs of the historical data base. The A-6 program and other airframe programs in the historical data base were developed and produced using the current state-of-the-art technology. The primary material was aluminum. The workers had experience working with the material. In addition, efficient manufacturing methods had been developed for working with aluminum. Therefore, material cost did not decrease much as follow-on lots were produced, resulting in a higher percentage learning-curve slope. The F-14 program used titanium and some carbon expoxy composites. Associated with these new materials are three factors which cause a lower percentage learning-curve to be reflected in the material cost category. The first factor is economics. As the supply of a new material increases to meet demand, the price of the new material decreases. The second factor results from increased learning by the worker to work with a new material. As the worker learns to work with a new material the material wastes are reduced resulting in decreased material cost. Related to the above factors are improvements in manufacturing techniques for producing and working the new material. The second reason is tolerance requirements. The F-14 airframe was required to perform at a higher speed and operating ceiling. Because of this increased demand on the F-14 airframe, closer material tolerances are required than for the A-6 airframe. Increased learning for the F-14 program may have resulted from reduced waste due to worker learning and improved manufacturing techniques to meet the closer tolerances. The last reason for the difference in quantity coefficients may be caused by having to estimate the acceptance schedule for the F-14 program. As previously discussed, the acceptance schedules were required to calculate cost per lot in constant year dollars. Errors due to conversion to constant year dollars directly affect the learning-curve since the learning-curve is a function of dollars versus quantity. Therefore, the F-14 program material cost cannot be used with confidence for determining the predictive capability of the Bayesian equations for the material cost category.

A final observation made analyzing the data in Table XX and seen in the data previously presented in this chapter is the weighting effect on the Bayesian methodology. As more information about the new program is considered, a higher probability exists that the prediction will be the true value. The weighting effect is illustrated by the analysis performed on the labor cost category. The prediction interval decreases or converges on the prediction of the next lot as additional lots are considered. For example, the upper limit percent deviation decreases from 84 percent for lot two to 78 percent for lot 5 for the A-6 program cumulative labor cost predictions. The same convergence is seen in the marginal cost predictions.

Table XXI presents the cumulative fifth lot hour/cost predictions using the various methods discussed previously and the percent deviation of the prediction from the actual lot cost. Using percent deviation as the measure for determining the best methodology, ordinary regression resulted in estimates having the smallest percent deviation for three out of four of the cost categories when four lots of information were available. For the material cost category, regression was more consistent across the two test programs. For example, the Levenson methodology had a .01 percent deviation for the F-14 program and a 68 percent deviation for the A-6 program. Ordinary regression resulted in percent deviations of -14 and 2 respectively for the F-14 and A-6 programs. Regression provided the smaller and more consistent percent deviation of prediction from the actual. However, ordinary regression applied to only four observations for predicting the next unit may

Table XXI
Cumulative Fifth Lot Hour/Cost Predictions
(in thousands)

Functional Area Method Used	A-6 Program		F-4 Program	
	Prediction	Percent Dev	Prediction	Percent Dev
ENGINEERING				
Levenson	6636	4	16698	-15
RANDOM	5582	-12	16255	-17
Bayesian	7304	15	21470	9
Regression	6232	-2	20268	3
TOOLING				
Levenson	12906+	208	11522	19
RANDOM	8974+	114	10756	11
Bayesian	4671	12	11011	13
Regression	4526	8	9692	-.1
LABOR				
Levenson	20749	20	31081	18
RANDOM	20945	21	30615	17
Bayesian	15042	-13	24929	-5
Regression	17880	3	26574	1
MATERIAL				
Levenson	127704+	68	229175	.01
RANDOM	78216	3	182876+	-20
Bayesian	67457	-11	328193	43
Regression	77841	2	195856+	-14

+Prediction not within Bayesian Equations 95% Prediction Interval have a high risk if the data has a large amount of variability.

Table XXII compares the various predictions of marginal fifth lot hour/cost for the A-6 and F-14 programs. Except for the engineering category, the Large equations developed to predict marginal cost were not as good as either the Bayesian or ordinary regression equations. Using the percent deviation as a measure, the Large equations were second to the Bayesian equation for the A-6 program and second to ordinary regression equation for the F-14 program. An explanation cannot be provided by the writer for the anomaly observed in this cost category.

For each program and cost category a methodology can be selected for providing the best prediction capability using the percent deviation from actual cost as a measure of "best". However, across both test programs and all categories the Bayesian methodology appeared more consistent, followed by ordinary regression, for providing the "best" prediction of future lot costs.

Tables XXIII and XXIV show the predictions for cumulative and marginal labor hours for future lots respectively. The analysis involved adding information about one more lot of airframes to predict the next lot. The addition of lot information was only accomplished for the Bayesian methodology and ordinary regression. With information on three lots, ordinary regression of cost versus quantity was performed.

Table XXII
Marginal Fifth Lot Hour/Cost Predictions

Functional Area Method Used	A-6 Program		F-14 Program	
	Prediction	Percent Dev	Prediction	Percent Dev
ENGINEERING				
Levenson	11032	28	36429	40
RANDOM	10856	26	36721	41
Bayesian	8737	-2	34439	32
Large	7376	-14	31527	21
Regression	4855	-43	21647	-17
TOOLING				
Levenson	45732+	432	30863	48
RANDOM	45770+	432	31589	51
Bayesian	10626	33	19628	-7
Large	25881	201	28260	35
Regression	7862	-9	12896	-38
LABOR				
Levenson	155506	51	180274	39
RANDOM	163937	59	181311	40
Bayesian	95185	-8	141484	10
Large	144980	40	178738	38
Regression	103714	.04	135064	4
MATERIAL				
Levenson	827385	62	1441495	13
RANDOM	761890	49	1324161	4
Bayesian	491227	-4	1362786	7
Large	811768	59	1300612	2
Regression	571446	12	655742	-48

+Prediction Not Within Bayesian Equations 95% Prediction Interval

However, the equation has only one degree of freedom. For one observation or two observations, regression could not be performed. When information for two lots were available, the equation for a straight line between two points was derived and used to predict the cost of the next lot(s).

The RANDOM equation was used in the analysis to serve as a baseline upon which no update was performed. The same RANDOM equation was used to predict lots 2, 3, 4, and 5. Separate Bayesian and regression equations were derived as each new observation was added to the data base. The resulting equations were used to predict the remaining future lots, up to the fifth lct.

Comparing the percent deviation of the prediciton from the actual the RANDOM equation does not provide the best estimate of cost of the next lot. The Bayesian, regression, or straight line equations are better predictors. The same result was observed for the labor cost category for four observations, Tables XXI and XXII.

For cumulative cost predictions, the straight line and regression methods result in lower percent deviations than obtained using Bayesian equations. However, in the case of marginal predictions the Bayesian methodology results in better prediction capabilities than straight line and regression equations to predict lot four. The regression equation becomes a better predictor of lot five labor costs. Because the regression equation has only one

Table XXIII
Cumulative Prediction of Labor Hours
For Future Lots Using 1, 2, and 3 Observations

Type CER	No. Obser.	AC Type	Predictions* Lot 2	Lot 3	Lot 4	Lot 5	Percent Deviation** Lot 2	Lot 3	Lot 4	Lot 5
Random	1	A-6	6740	10753	15707	20926	48	35	27	21
	1	F-14	7013	14173	23333	30586	30	24	16	17
Bayesian	1	A-6	3931	6293	9218	12306	-14	-21	-26	-29
	1	F-14	5348	10848	17904	23502	-.7	-5	-11	-10
Random	2	A-6		10753	15707	20926		35	27	21
	2	F-14		14173	23333	30586		24	16	17
Bayesian	2	A-6		6806	9994	13366		-15	-19	-23
	2	F-14		10885	17967	23587		-5	-10	-10
Straight	2	A-6		8365	13673	19834		5	10	14
Line	2	F-14		11253	18973	25196		-2	-5	-4
Random	3	A-6			15707	20926			27	21
	3	F-14			23333	30586			16	17
Bayesian	3	A-6			10650	14295			-14	-18
	3	F-14			18358	24131			-8	-8
Regression	3	A-6			12955	18558			5	7
	3	F-14			19415	25893			-3	-1

*In Thousands **From Actual Cost

Table XXIV
Marginal Prediction of Labor Hours
For Future Lots Using 1, 2, and 3 Observations

Type CER	No. Obser.	AC Type	Predictions Lot 2	Lot 3	Lot 4	Lot 5	Percent Deviation ** Lot 2	Lot 3	Lot 4	Lot 5
Random	1	A-6	362025	263712	202245	163980	85	77	70	59
	1	F-14	482129	331734	229969	181361	50	42	28	40
Bayesian	1	A-6	201911	149567	116327	95371	3	.5	-2	-8
	1	F-14	407069	283537	198929	158107	27	22	11	22
Random	2	A-6		263712	202245	163980		77	70	59
	2	F-14		331734	229969	181361		42	28	40
Bayesian	2	A-6		147728	114879	94172		-.6	-3	-9
	2	F-14		260868	183335	145872		12	2	13
Straight	2	A-6		160591	135792	118933		8	14	15
Line	2	F-14		202258	128650	95954		-13	-28	-26
Random	3	A-6			202245	163980			70	59
	3	F-14			229969	181361			28	40
Bayesian	3	A-6			115126	94382			-3	-9
	3	F-14			177620	141204			-.1	9
Regression	3	A-6			125574	108153			6	5
	3	F-14			152608	118223			-15	-9

**From Actual Cost

degree of freedom, may explain why regression was not better than the Bayesian method for predicting lot four.

Figure 6, shows two plots of the results of the analysis of the labor category. The top graph displays the cumulative labor hour predictions for the A-6 program. The bottom graph displays the cumulative labor hour predictions for the F-14 program. Both graphs portray identical trends, therefore, the discussion will only address the A-6 program predictions. The slope of the line drawn through the prediction values from the RANDOM equation represents an average for cumulative labor data for fighter type aircraft since the RANDOM equation was derived using historical fighter aircraft data. The line of plotted actual values represents the learning-curve slope unique to the A-6 program. Straight line and ordinary regression yield point estimates as new data are added. The straight line and regression equation point estimates show relatively small deviations from the actual values because the actual values show very little variance.

Although the Bayesian estimates appear to have a constant slope the points cannot be validly connected because the estimates are also point estimates. The coefficient of the quantity term (indication of learning-curve slope) and the adjustment term of the Bayesian equation changes with addition of new information. Therefore, the predictions using the Bayesian equations are point estimates.

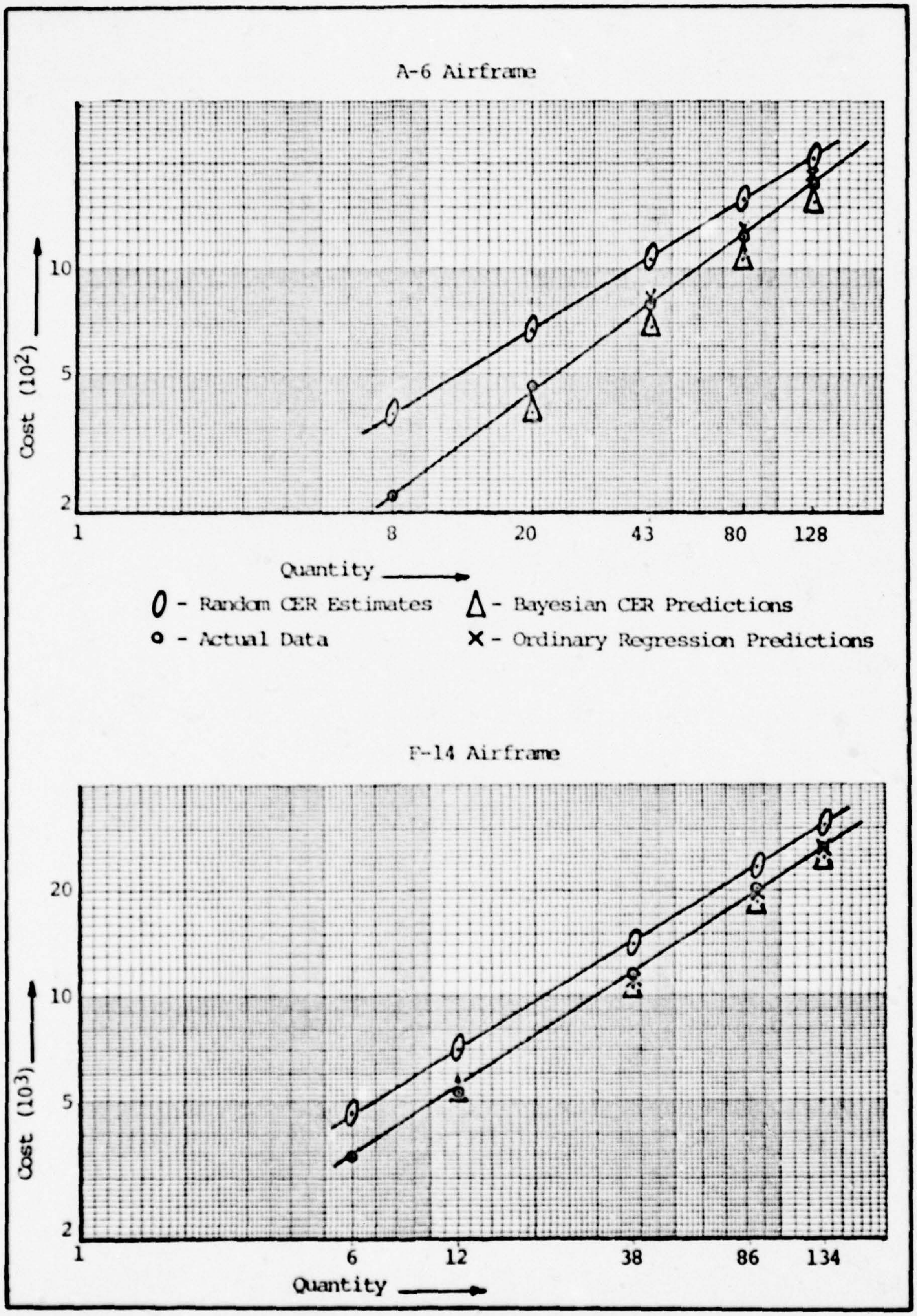


Figure 6. Cumulative Recurring Labor Hour Data and Predictions

Although the Bayesian prediction understates the lot cost when compared with the actual cost, the Bayesian estimate is closer to the actual value than the estimate using the RANDOM equation.

Figure 7 shows the comparison of future lot marginal labor hour cost predictions using first, second, third and fourth lot costs for both the A-6 and F-14 program. The results observed were basically the same as discussed for Figure 6, except the straight line and regression predictions showed more deviation from the actual values. This may be due to variance being introduced in the data as a result of the estimation of the learning-curve slope for calculation of true lot midpoints. The Bayesian methodology uses the variance between observed lots in development of the equation. The variance is used to estimate an adjustment factor for the prediction equation.

Summary

In this chapter, the marginal and cumulative CERs developed using the Bayesian methodology were presented for the cost categories of total engineering, total tooling, recurring labor, and recurring material. Using the coefficients of the explanatory variables, t-statistics, variance and standard deviation of prediction, the Bayesian CERs were compared with the RANDOM CERs and CERs developed using ordinary regression. The Bayesian CERs were

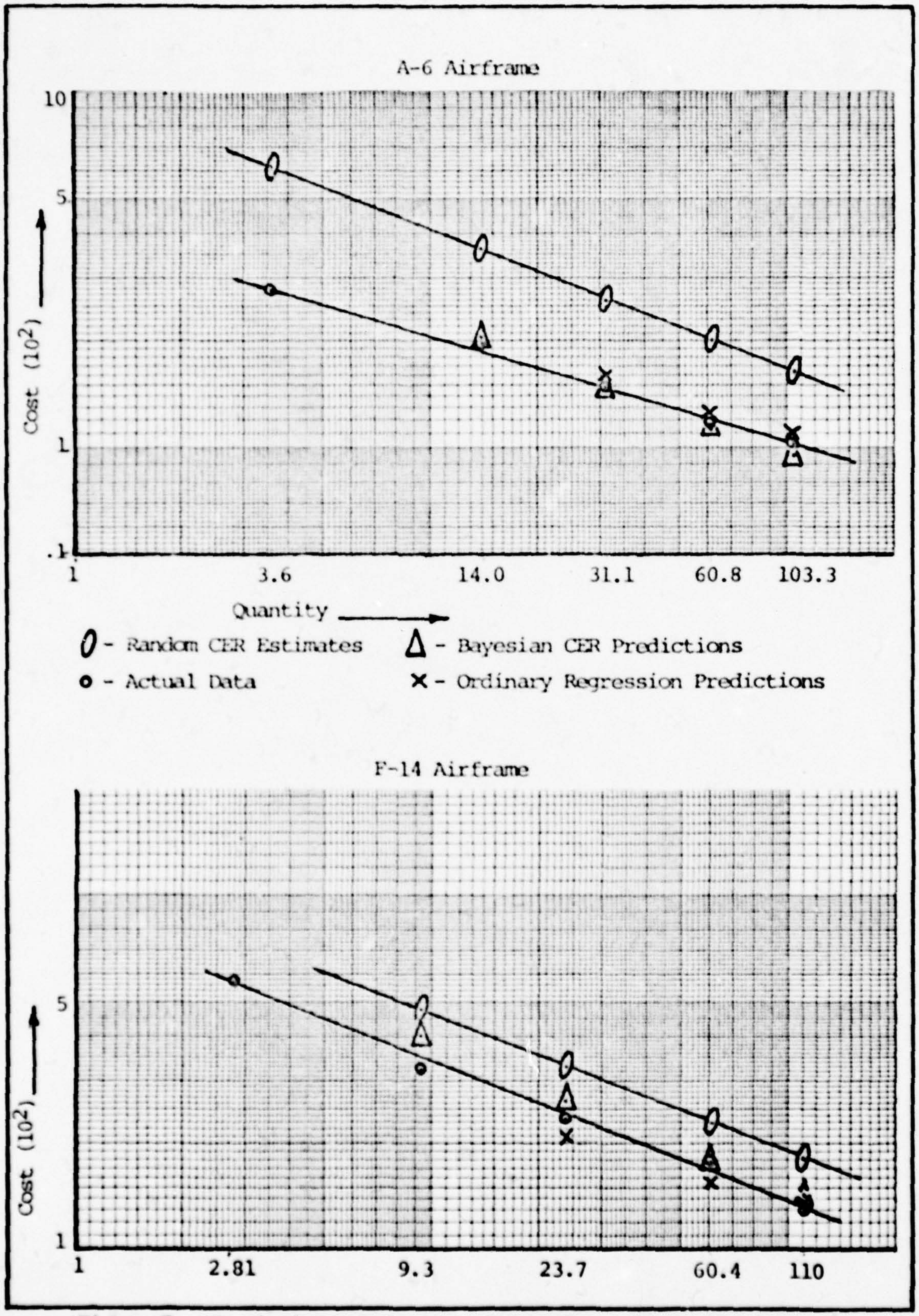


Figure 7. Marginal Recurring Labor Hour Data and Predictions

developed using historical lot cost data and cost information on the first four lots of a test case airframe program. Additional Bayesian equations and straight line or regression equations were developed for the category of labor. These equations used the first, second, and third lot cost information for predicting the cost of future lots.

The chapter also presented the fifth lot cost predictions obtained using ordinary regression, Bayesian, RANDOM, Levenson, and Large equations for the A-6 and F-14 programs. The percent deviation of the prediction from the actual cost was used as the measurement for comparison of the methodologies.

Predictions of the next lot cost using one, two, or three observations and ordinary regression, Bayesian, and RANDOM equations were compared for the labor category. The labor category predictions were graphically displayed. Predictions and upper and lower limits of a 95% confidence interval were calculated for each Bayesian CER. The next chapter presents the conclusions drawn from the results and comparisons presented in this chapter.

VI. Summary and Conclusion

As a result of high inflation and a smaller (percentage of DOD budget) approved budget for research and development of new weapons systems, increased emphasis has been placed on accurate cost estimates. Cost estimates are required for making intelligent decisions concerning expenditure of budget dollars.

DSARC III is a DOD review committee chaired by the Secretary of Defense which makes the production decision for acquisition of major weapon systems. Because this decision commits a large amount of budgetary resources, a reliable cost estimate is essential. Program Managers require reliable cost estimates to budget for future production lot acquisitions and control future expenditures. Currently, production phase estimates are provided by applying learning-curve theory to "grass roots" cost data obtained during the prototype phase or from earlier lot productions. These estimates may be in error due to the inaccuracies in the cost reporting system or biases in the selection of a learning-curve slope.

The objective of this thesis was to investigate a statistical technique to provide a reliable estimate of the cost of the next lot using the available early actual cost data. The statistical technique developed should provide a quick independent method for evaluation of estimates obtained using current techniques and methodologies. The nature of the problem to estimate the cost of the next lot or unit using early actual information lead to the selection of a Bayesian approach for updating a CER to predict the cost of the next lot of airframes.

Fighter airframe cost data was used for several reasons. First, airframe cost data has exhibitied a direct casual relationship with the explanatory variables (CERs). Secondly, data adjusted for consistency was readily available from a recent RAND study (Ref 17). Next, limiting the study to one type of airframe provides a more homogeneous data base. Finally, the fighter airframe subset of data consisted of a large number of observations for performing statistical analyses.

The Bayesian approach was applied to the RANDOM equations to obtain updated airframe CERs. The RANDOM equations were chosen because they prove to be statistically better predictors than equations developed using other recently published methodologies. The RANDOM equations were developed from a mixed linear model which considers two sources of error. The first source of error is due to different types of airframes. The second source of error is the overall regression error.

GLS techniques were required to estimate the various parameters of the Bayesian CER and the RANDOM equations. The primary reason for using GLS was the lot observations within an airframe program were assumed to be correlated. For example, lot two shares dependency with lot one in the areas of design, manufacturing labor, etc.

Based on the definition of recurring costs, production costs and recurring costs are directly related. Because of this relationship, a production cost estimate can be obtained using recurring cost information. Therefore, development of the Bayesian equations to predict the production cost of the next unit for the labor and material cost categories used recurring costs for the first, second, third, and fourth units of the new program and the recurring cost of the historical lot observations. For the engineering and tooling categories the separation of costs into recurring and nonrecurring costs was difficult. As discussed in Chapter III, inconsistencies exists because of definitional problems. Due to the inconsistencies, total cost was used to develop the Bayesian equations for the engineering and tooling categories.

Several comparisons were presented in Chapter V. Four lot observations of each test case were used to develop the ordinary regression equations and Bayesian equations. The coefficient of the explanatory variables of the Bayesian equations were compared with RANDOM and ordinary regression equations. The t-statistic values for the quantity variable indicate that this variable is significant for predicting the lot cost of fighter airframes. The value of variance for the Bayesian equations are less than the values of variance for the RANDOM equations. The smaller value of variance should be expected as more new information becomes available because Bayesian philosophy allows consideration of all knowledge, prior and current, and then weighing the information accordingly. The Bayesian predictions using four lots of information have a lower percentage deviation from the actual cost than the RANDOM equation prediction.

The labor hour category was investigated using one, two, or three lot observation in the development of the Bayesian equation. The predictions obtained from the Bayesian equations were consistently better than the predictions using the RANDOM equations. The greater the lot number being predicted the better the prediction using both the RANDOM and Bayesian equations. This phenomenon can be observed in Table XX. The percent deviation for the next lot prediction is less than the percent deviation for the prediction of the previous lot.

Figures 6 and 7 illustrate the convergence of the prediction and actual value for the analysis of the labor category. For the Bayesian prediction this result can be explained. As additional cost information is used in the Bayesian approach, the new information is weighted and adjusts the resultant CERs such that the prediction of the next lot is closer to the actual value. The convergence of the RANDOM estimates to the actual values cannot be explained by this writer.

Using the marginal fifth lot hour/cost predictions, for both the A-6 and F-14, Table XXII, and the percentage deviation as a measurement of predictability, the Bayesian approach, except for ordinary regression, resulted in predictions closer to the actual cost than did the other methodologies presented. Ordinary regression equations must be regarded with caution since these equations were developed using only four lots of information and have three degrees of freedom. With any variability in the data predictions made using the regression equations would be unreliable. Therefore, the Bayesian approach results in equations which consistently provide reliable cost predictions of the next lot. This conclusion is further supported by the detailed analysis and comparisons performed on the labor cost category.

Recommendations

This thesis effort was an extension of the study, Aircraft Airframe Estimation Utilizing a Components of Variance Model by Marcotte (Ref 19). The underlying assumption of the study by Marcotte was, "there is correlation between observations of different lots of the same type airframe (Ref 19:6)." Therefore, this assumption underlies the Bayesian approach used in this effort.

Based on the above information and assumption, an obvious effort requiring investigation is the application of GLS, instead of ordinary regression, to prediction of the cost of the fourth and fifth lots of a new program using quantity as the only explanatory variable. Applying GLS using three or more observations of recurring cost may provide a better estimate of production cost since GLS considers correlation and adjusts for variability of the data.

Another area requiring further investigation is treatment of the learning-curve as a random effect. Because the learning-curve is significantly affected by the quantity term and lot quantity purchases are random, the quantity explanatory variable should be treated as a random effect in Henderson's Method 3 mixed linear model. This recommendation is supported by Tables VIII-XV. The coefficient of the quantity variable (an indicator of learning-curve slope) for the RANDOM equations represents an average learning-curve for all lot observations and programs. The coefficient of the quantity term of the ordinary regression equation developed using data for a test case program represents a unique learning-curve slope for that program. Both the A-6 and F-14 programs have learning-curves which are different from the average learning-curve slopes of the historical data base. The closer the particular cost category learning-curve for a specific airframe program (A-6 or F-14) was to the average learning-curve slope resulted in a better prediction of the cost of the next lot using the Bayesian equation.

Furthermore, the coefficient of the quantity term of the RANDOM equation changed to a value in the Bayesian equation closer to the value reflected in the ordinary regression equation.

An observation, undoubtedly an obvious one, is the accuracy of future program cost estimates for a new program increases directly with additional knowledge of the "true" learning-curve for the program.

Bibliography

1. AFLCP/AFSCP 800-15. Acquisition Management Contractor Cost Data Reporting (CCDR). Washington: Department of the Navy, Army, and Air Force, November 1973.

2. AFR 800-2. Acquisition Management Program Management. Washington: Department of the Air Force, March 1972.

3. AFSCP 800-3. Acquisition Management A Guide for Program Management. Washington: Department of the Air Force, May 1971.

4. An Orientation to Air Force Acquisition Management and the System Program Office. Briefing. Deputy for Systems Aeronautical Systems Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio.

5. Asher, Harold. Cost-Quantity Relationships in the Airframe Industry. The RAND Corporation Research Memorandum R-291. Santa Monica, California: The RAND Corporation, July, 1956.

6. Bannon, P. W. An Analysis of the Learning-Curve in Relation to Modern Fighter Production. Unpublished thesis. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, December 1967.

7. Boren, H. E., Jr., and H. G. Campbell. Learning-Curve Tables, Volume I, II and III. RAND Corporation Memorandum RM-6191-PR. Santa Monica, California: The RAND Corporation, April 1970.

8. Chase, R. B., and N. J. Aquilano. Production and Operations Management. Homewood, Illinois: Richard D. Irwin, Inc., 1973.

9. Fabrycky, W. J., and P. E. Torgersen. Operations Economy: Industrial Applications of Operations Research. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

10. Fisher, G. H. Cost Considerations in Systems Analysis. The RAND Corporation Report R-490-ASD. Santa Monica, California: The RAND Corporation, 1971.

11. Freund, J.E. Mathematical Statistics. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1971.

12. Goldman, T. A. Cost-Effectiveness Analysis: New Approaches in Decision Making. New York: Fredrick A. Praeyer, Publishers, 1967.

13. Handel, V. Aircraft Airframe Cost Estimation by the Application of Joint Generalized Least Squares. Unpublished thesis. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, November 1975.

14. Johnston, J. Econometric Methods. New York: McGraw-Hill Book Company, Inc., 1963.

15. Johnston, J. Statistical Cost Analysis. New York: McGraw-Hill Book Company, Inc., 1960.

16. King, W. R. Probability for Management Decisions. New York: John Wiley & Sons, Inc., 1968.

17. Large, J. P., H. G. Campbell and D. Gates. Parametric Equations for Estimating Aircraft Airframe Costs. The RAND Corporation Report R-1693-PA&E. Santa Monica, California: The RAND Corporation, May 1975.

18. Levenson, G. S., and H. E. Boren, Jr., et al. Cost-Estimating Relationships for Aircraft Airframes. The RAND Corporation Report R-761-PR. Santa Monica, California: The RAND Corporation, 1971.

19. Marcotte, R. C. Aircraft Airframe Cost Estimation Utilizing a Components of Variance Model. Unpublished thesis. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, October 1976.

20. McDonald, P. R. The Theory and Practical Application of Improvement Curves. Corvina, California: Procurement Associates, 1967.

21. Military Equipment Cost Analysis. Santa Monica, California: The RAND Corporation, 1971.

22. Morgan, B. W. An Introduction to Bayesian Statistical Decision Processes. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1968.

23. Morris, W. T. The Analysis of Management Decisions. Homewood, Illinois: Richard D. Irwin, Inc., 1964.

24. Novick, D. Are Cost Overruns a Military-Industry-Complex Specialty? RAND Corporation Pamphlet P-4311. Santa Monica, California: The RAND Corporation, March 1970.

25. Pichon, A. A., Jr. The Development of a Predictive Model for First Unit Costs Following Breaks in Production. Unpublished masters' thesis. Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, August 1974.

26. Puryear, V. A. New Progress Functions with Probability Modifiers. St. Louis: Office of the Comptroller Cost Analysis Division, U.S. Army Troop Support Command, May 1973.

27. Radford, K. J. Managerial Decision Making. Reston, Virginia: Reston Publishing Company, Inc., 1975.

28. Raiffa, H., and R. Schlaifer. Applied Statistical Decision Theory. Cambridge, Massachusetts: The MIT Press, 1961.

29. Sanchez, L. R., and J. A. Deirossi. Methods of Estimating Fixed-Wing Airframe Costs. Planning Research Corporation Report PRC-R-547A, Vol I. Los Angeles: The Planning Research Corporation, 1967.

30. Sasaki, K. Statistics for Modern Business Decision Making. Belmont, California: Wadsworth Publishing Company, Inc., 1969.

31. Scott, L. S. A New Approach to Weapon System Cost Estimating. Masters' thesis. Monterey, California: Naval Postgraduate School, October 1966.

32. Searle, S. R. Linear Models. New York: John Wiley & Sons, Inc., 1971.

33. Summers, R. Cost Estimates as Predictors of Actual Weapons Costs: A Study of Major Hardware Articles. RAND Corporation Memorandum RM-3061-PR. Santa Monica, California: The RAND Corporation, March 1965.

34. Theil, H. Principles of Econometrics. New York, New York: John Wiley & Sons, Inc, 1971.

35. Trueger, P. M. Accounting Guide for Defense Contracts. New York: Commerce Clearing House, Inc., 1966.

36. Womer, N. K. Cost Analysis and Applied Statistical Inference. Unpublished paper. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, 1975.

37. Wright, T. P. Factors Affecting the Cost of Airplanes. Journal of the Aeronautical Sciences, Vol. 3, 122-128. February 1936.

38. Yates, E. H. A Method for Estimating the Procurement Cost of Aircraft. General Electric TEMPO Report RM 60 TMP-39. Santa Barbara, California: General Electric Company, 1969.

Data Source

RAND Aircraft Airframe Cost Data Worksheets. Unpublished historical cost summary. Santa Monica, California: The RAND Corporation.

Supplementary Bibliography

Alchian, A. A. Reliability of Cost Estimates-Some Evidence. RAND Corporation Memorandum RM-481. Santa Monica, California: The RAND Corporation, October 1950.

Box, G. E. P. and G. C. Tiao. "A Bayesian Approach to the Importance of Assumptions Applied to the Comparison of Variances," Biometrika, Vol. 51 No. 1 and 2: 153-167 (1964).

Brown, R. V. Research and the Credibility of Estimates: An Appraisal Tool for Executives and Researchers. Boston: Division of Research Graduate School of Business Administration Harvard University, 1969.

Colasuonno, V. An Analysis of Progress Curve Conceptual Advances and Progress Curve Uses, Since 1956. Unpublished thesis, Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, September 1967.

DOD LCC-3. Life Cycle Costing Guide for System Acquisitions. Washington: U.S. Government Printing Office, January 1973.

Dunne, W. E. Microeconomic Theory Applied to Parametric Cost Estimation of Aircraft Airframes. Unpublished thesis. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, December 1975.

Dyckman, T. R., S. Smidt, and A. K. McAdam. Management Decision Making Under Uncertainty. London: The Macmillian Company, 1969.

Efron, B., and C. Morris. Limiting the Risk of Bayes and Empirical Bayes Estimators, Part II. The Empirical Bayes Case. RAND Corporation Report R-647-PR. Santa Monica, California: The RAND Corporation, June 1971.

Fisher, G. H. Derivation of Estimating Relationships: An Illustrative Example. RAND Corporation Memorandum RM-3366-PR. Santa Monica, California: The RAND Corporation, November 1962.

Heschel, M. S. Bayesian Decision Theory. Unpublished masters' thesis. Columbus, Ohio: The Ohio State University, 1965.

Large, J. P., K. Hoffmayer, and F. Kontrovich. Production Rate and Production Cost. RAND Corporation Report R-1609-PA&E. Santa Monica, California: The RAND Corporation, December 1974.

Lindgren, B. W. Elements of Decision Theory. New York: The Macmillian Compnay, 1971.

Luce, R. D., and H. Raiffa. Games and Decisions. New York: John Wiley & Sons, Inc., 1958. 2nd printing.

Morris, W. T. Management Science: A Bayesian Introduction. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1968.

Newman, J. W. Management Applications of Decision Theory. New York: Harper & Row Publishers, 1971.

Noah, J. W., and J. M. Daniels, et al. Estimating Aircraft Acquisition Costs by Parametric Methods. United States Navy Report FR-103-USN, Alexandria, Virginia: J. Watson Noah Associates, Inc., 1973. AD 913440

Pratt, J. W., H. Raiffa, and R. Schlaifer. Introduction to Statistical Decision Theory. New York: McGraw-Hill Book Company, 1965.

Press, J. S. Statistical Estimation by Empirical Bayes Method: Some Extensions and Logistical Applications. RAND Corporation Memorandum RM-4442-PR. Santa Monica, California: The RAND Corporation, June 1965.

Radtummala, V. M. Decision Analysis with Business Applications. New York: Intext Educational Publishers, 1973.

Roscoe, J. T. Fundamental Research Statistics for the Behavioral Sciences. New York: Holt, Rinehart, and Winston, Inc., 1969.

Schmitt, S. A. Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics. Reading, Massachusetts: Addison-Wesley Publishing Company, 1969.

Tiao, G. C. and A. Zellner. "Bayes' Theorem and the Use of Prior Knowledge in Regression Analysis." Biometrika, Vol. 51: 219-230 (1964).

Timson, F. S., and D. P. Tihansky. Confidence in Estimated Airframe Cost: Uncertainty Assessment in Aggregate Predictions. The RAND Corporation Report R-903-PR. Santa Monica, California: The RAND Corporation, 1972. AD 757503

Wheelwright, S. C. Forecasting Methods for Management. New York: John Wiley & Sons, Inc., 1973.

Winston, W. A. A Comparison of Air Force and Bayesian Demand Rate Estimates Applied to Initial Provisioning of the C-141 Weapon System Program. Unpublished masters' thesis. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, August 1965.

Vita

Walter Dietrich was born on 27 August 1946 in Brunswick, Georgia. He graduated from high school in Belleville, Illinois in 1964 and then attended Belleville Junior College. He entered the University of Missouri at Rolla in 1966 from which he received a Bachelor of Science degree in January 1969. He completed pilot training in March of 1971 and became a C-130 co-pilot at Homestead AFB, Florida. He was assigned to Davis-Monthan AFB, Arizona as a DC-130 pilot prior to entering the Air Force Institute of Technology in September 1975.

Permanent address: 705B Hamel Drive
Lee Summit, Missouri 64063