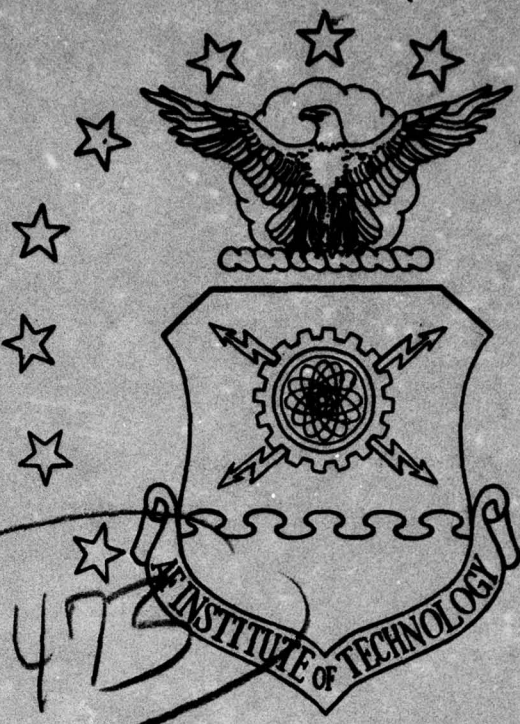
AD-A032 627 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO SCH--ETC F/G 5/1
AIRCRAFT AIRFRAME COST ESTIMATION UTILIZING A COMPONENTS OF VAR--ETC(U)
OCT 76 R C MARCOTTE

UNCLASSIFIED GOR/SM/76D-10 NL

1 OF 2
AD
A032627



AD A032627



DDC
RECEIVED
NOV 26 1976
B

UNITED STATES AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

ACCESSION for
NTIS White Section
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. AVAIL. and/or SPECIAL
A

AIRCRAFT AIRFRAME COST ESTIMATION UTILIZING A COMPONENTS OF VARIANCE MODEL

THESIS

GOR/SM/76D-10

Ronald Marcotte
Capt USAF

DDC
RECEIVED
NOV 26 1976
B

Approved for public release; distribution unlimited.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

GOR/SM/76D-10

AIRCRAFT AIRFRAME COST ESTIMATION UTILIZING A COMPONENTS OF VARIANCE MODEL

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

by

Ronald C. Marcotte, B.S.

Capt USAF

Graduate Operations Research

October 1976

Approved for public release; distribution unlimited.

Preface

Data correlation has not been explicitly addressed in any past studies which develop aircraft airframe cost estimating relationships. The Components of Variance model can recognize multiple sources of error and when adapted to airframe cost estimation can explicitly deal with this correlation problem and improve the predictive qualities of the resulting cost estimating relationship. The Introduction, Chapter I, and Conclusions, Chapter VI, outline the thrust and results of this thesis in a non-mathematical manner. The development and explanation of the techniques employed along with the specific results obtained are presented in the remaining chapters.

My sincere thanks and deep appreciation must be extended to Dr. N. Keith Womer, my thesis advisor, for his patience, support, and assistance in this effort. Also, without the great understanding and sacrifice of my wife, Ruth, this entire study would not have been possible.

Contents

		Page
Preface		ii
List of Tables		v
List of Figures		vii
Abstract		viii
I.	Introduction	1
	Airframe Costs	2
	Problems with Parametric Estimation	2
	Data Problems	3
	Hypotheses	6
	Application of Hypotheses	6
	Procedure	7
	Organization	8
II.	Components of Variance Model	9
	Notation	9
	Ordinary Least Squares	9
	Generalized Least Squares	11
	Extension of GLS to Airframe Cost Estimation	12
	Estimation of the Variance Components	15
	Outline of Variance Component Estimation Technique	16
	Relationship of the RANDOM Technique to the Techniques of Timson and Tihansky and Large, et al.	18
	A Possible Test of Hypothesis	19
III.	The CER's to be Developed and the Data	20
	Physical Characteristics (Explanatory Variables) Related to Cost	20
	Functional CER Forms	21
	Data Source	23
	Data Reduction	23
	Cost Element Definitions	24
	The CER's	25
	Cumulative Cost Approach	25
	Marginal Cost Approach	25
	LAR Technique/Cost Approach	29
	Summary	30
IV.	CER Development, Presentation, and Analysis	35
	CER Values to be Presented	35
	Total Engineering Hours	38
	Total Tooling Hours	41

Contents

		Page
	Recurring Labor Hours	43
	Recurring Material Dollars	45
	Comparisons Overall	45
V.	Prediction Analysis	48
	F-14 Explanatory Variables	48
	Adjusting Lot Two Predictions	49
	Prediction Intervals	52
	Prediction Presentation	56
VI.	Summary and Conclusions	64
	New Research	67
Bibliography		69
Data Source		72
Supplementary Bibliography		74
Appendix A:	The Development of the Variance Component Estimation Procedure	76
Vita		87
Appendix B:	Computer Program (Separate Cover)	88
Appendix C:	Data (Separate Cover)	95
Appendix D:	Actual Versus Predicted Costs for F-14 (Separate Cover)	99

List of Tables

Table		Page
I	Price Adjustment Indices	26
II	Aircraft Characteristics (Explanatory Variables)	32
III	Remaining Quantity Values for the Marginal Cost Approach	34
IV	Definitions	37
V	CER Coefficients and CER Statistics Engineering Hours	39
VI	CER Coefficients and CER Statistics Tooling Hours	42
VII	CER Coefficients and CER Statistics Labor Hours	44
VIII	CER Coefficients and CER Statistics Material Dollars	46
IX	Explanatory Variable Values used for Prediction	49
X	Engineering Hour Predictions	57
XI	Tooling Hour Predictions	58
XII	Labor Hour Predictions	59
XIII	Material Dollar Predictions	60
XIV	Average Percent Intervals	62
XV	Program Variables Explained	90
XVI	Acquisition Cost Data - Cumulative	96
XVII	Acquisition Cost Data for 100 Airframes	97
XVIII	Explanatory Variables	98
XIX	Engineering Hours - Actual Versus Predicted	101
XX	Tooling Hours - Actual Versus Predicted	102
XXI	Labor Hours - Actual Versus Predicted	103

List of Tables (continued)

Table		Page
XXII	Material Dollars - Actual Versus Predicted	104
XXIII	Average Percent Deviation	105

List of Figures

Figure		Page
1	Standardized Residuals Versus Weight	5

Abstract

Previous studies into airframe acquisition cost estimation do not explicitly recognize the existence of correlation in the historical data. If one believes this data problem exists, then it is possible to develop a components of variance model that takes the problem into account. It is a more general model that recognizes two sources of error: (1) error due to different types of airframes and (2) overall or ordinary regression error. The variance of these two errors can be estimated and then can be utilized along with the technique of generalized least squares to obtain a cost estimating relationship which explicitly accounts for the data correlation. This modeling technique, when compared to techniques presently in service, shows that present estimating relationships underestimate the variance of the cost prediction of a new type airframe and overestimate the variance of the cost prediction of a follow-on airframe. Also, those existing techniques which implicitly recognize data correlation do not make use of all the data information available and therefore produce estimates with poor confidence/prediction intervals. The modeling technique developed here is an improvement over the present techniques utilized and advances the state of the art of parametric airframe cost estimation greatly.

AIRCRAFT AIRFRAME COST ESTIMATION UTILIZING

COMPONENTS OF VARIANCE MODEL

I. Introduction

It is becoming more and more important in industry and especially in the Department of Defense (DoD) to obtain good cost estimates of newly conceptualized products and weapon systems. DoD has been under ever increasing pressure from Congress and the American public to justify the choices and costs of new systems.

One of the more costly items that DoD buys is aircraft. Aircraft contracts run into billions of dollars and good cost estimation is essential when attempting to obtain Congressional approval for an aircraft program.

In the last ten years the parametric statistical approach to developing cost estimating relationships (CER's) to predict costs has received a great deal of emphasis and use. This is especially true in the case of predicting aircraft airframe acquisition costs. This paper will only address this particular area of parametric cost estimation, that is, aircraft airframe acquisition costs. An airframe is the body/frame of an aircraft without engines, avionics, wheels, etc. An aircraft contract is the agreement with the company picked to produce the aircraft. A "lot" of aircraft is what is delivered to DoD and this is ordinarily one year's buy.

Airframe Costs

There are many approaches to parametric airframe cost estimation. (Throughout the remainder of this study the term cost will refer to acquisition cost only and airframe acquisition cost will be the only item discussed.) The cost to be estimated could be the total cost of an airframe contract (lot). Conversely, the costs to be addressed could be the separate cost elements included in the total lot cost, i.e., tooling cost, material cost, labor cost, engineering cost, etc. The elemental cost approach necessitates the development of a CER for each cost element. Then to estimate the total lot cost, the cost of each element is estimated and these are added together.

A recent study by RAND Corporation reports that neither the total lot cost nor elemental lot cost approach results in superior accuracy (Ref 13:3-6). Therefore, because the elemental cost approach is able to point out to DoD the more expensive portions of cost, it is the approach most often utilized. This report will also use the elemental approach for the reasons given above and to facilitate comparison with prior work in this area of cost estimation.

Problems with Parametric Estimation

The RAND Corporation has been the pacesetter in the area of parametric airframe cost estimation. The different reports have approached the CER development in several different ways and they have studied many of the associated problems.

One problem is the choice of the functional form to be utilized for the CER model. This form could be linear, exponential, or log-linear, to name a few. Obviously, the specific choice of a functional form can change the results greatly.

Another problem area is the choice of a statistical technique. After the functional form is chosen one must use a specific technique with its inherent assumptions to estimate the CER's.

The choice of which independent variables (explanatory variables) to use is yet another area that causes problems. In airframe cost estimation the number of possible variables is enormous (from airframe unit weight to the number of wheels, etc.).

The final problem associated with parametric cost estimation to be mentioned in this report is the treatment of the data to be utilized. One can aggregate it, separate it, throw some out, make some up, and so on. The list of possible treatments is endless.

This thesis will ignore all the problems mentioned above except the last, data treatment. The functional form of all the CER's developed will be log-linear and the only explanatory variables to be utilized will be the natural logarithm of maximum aircraft speed at best altitude (S), unit airframe weight (W) and airframe quantity (Q).

Data Problems

There are basically two ways of utilizing historical airframe costs to develop cost estimating relationships. The first method is to treat every airframe contract as an observation. Of course, no entire inventory of a specific type aircraft is produced as one contract. Normally, there are several lots of an aircraft successively produced by the same contractor. For example, two recent RAND reports, one by Levenson, et al. and the other by Timson and Tihansky use this approach with 124 observations (lots). This encompasses only 26 different types of aircraft (Ref 10; Ref 13)!

The other approach to data utilization most often used is at the opposite end of the spectrum from the first. One observation per aircraft type is constructed (estimated) from the available contract data.

This data point estimation is most often accomplished by associating a cost with each aircraft (airframe) type at some normalised and/or specified airframe quantity. The cost of this given quantity may be difficult, if not impossible, to obtain from the contractors so most often it is derived by estimating the learning curve associated with each airframe type and then this curve is used to estimate the cost of, for example, 100 airframes. Of course, a curve must be estimated for each different aircraft type. RAND has studied the problem in this way several times in the past and two representative reports are by Levenson & Barro and Large, et al. (Ref 8 ; Ref 9).

There are problems associated with each approach as outlined above. There could be a great deal of correlation between observations under the lot cost approach. Womer suggests the possibility that lot costs of the same type airframe are highly correlated to one another when compared to lot costs of different or new airframes. Womer goes on to illustrate this point by graphing the standardized residuals obtained from a CER developed by Handel (Ref 4) against one of the explanatory variables. The CER was based on 37 observations of contracts for military fighter/trainer type airframes of 10 different types and Handel stated that this relation has a coefficient of determination of .94. Fig. 1 is a reproduction of this graph constructed by Womer and one can see that it suggests that there is strong evidence of this type of correlation (Ref 16:10-12).

The other data treatment would definitely be free of the correlation problem previously mentioned, but notice that the number of observations would be drastically reduced. For example, the Timson and

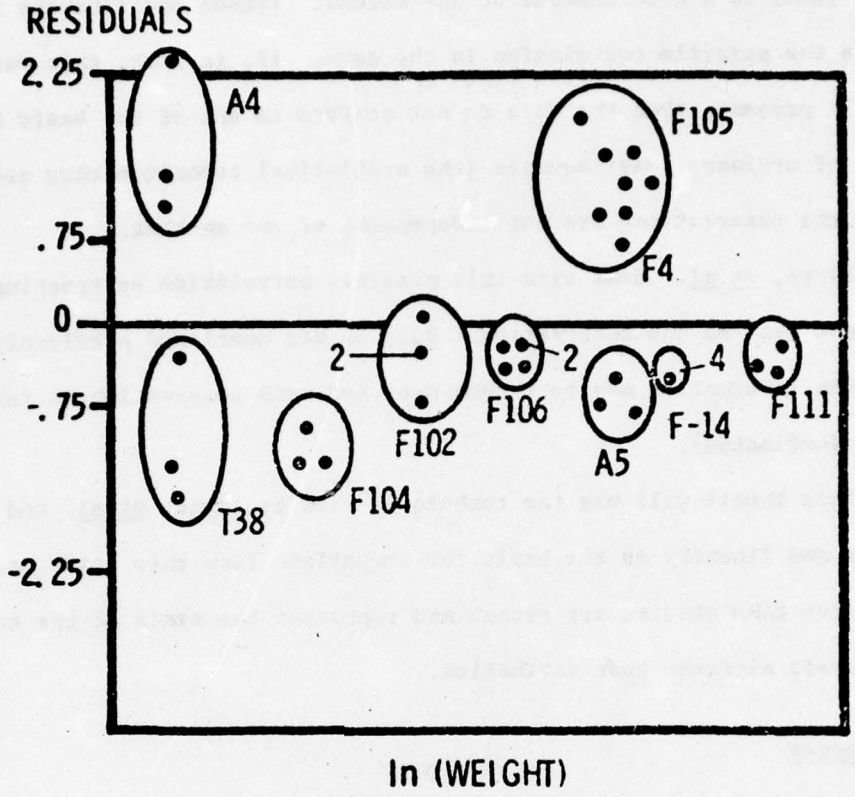


Fig. 1.* Standardized Residuals Versus Weight (From Ref 16:12)

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

Tihansky study used 124 observations whereas the Large, et al. report used only 26 (26 different type airframes). This could result in the loss of potentially valuable historical cost information. Also, one must remember that in the Large study the observations used were actually estimated or fictitious costs.

One can see that each of the data approaches has its associated problems. The Timson and Tihansky (TT) study is a very good representative study of the first treatment and the report by Large, Campbell and

Cates (LAR) is a good example of the second. Timson and Tihansky simply ignore the possible correlation in the data. If, in fact, this correlation is present, then the data do not conform to one of the basic assumptions of ordinary least squares (the statistical technique they used), i.e., the observations are not independent of one another.

Large, et al. deal with this possible correlation by treating each airframe type as one observation. But, as was mentioned previously, valuable information may be thrown away and each observation is fabricated (estimated).

This thesis will use the techniques used by Large, et al. and Timson and Tihansky as the basis for comparison from this point on. These two RAND studies are recent and represent the state of the art in parametric airframe cost estimation.

Hypotheses

The hypotheses of this study are that there is correlation between observations of different lots of the same type airframe and this correlation will be assumed to be caused by the existence of two sources of error associated with airframe costs. One source of error is due to different types of airframes and the second is due to overall data error (ordinary regression error).

A component of variance model is formulated to account for the two sources of error. This model is shown to be more general than both the linear models assumed by LAR and TT.

Application of the Hypotheses

If the hypotheses and assumptions given above are true, then present cost estimating relationships which use TT's approach will underestimate

the variance of predicting the cost of a totally new type of aircraft airframe. Conversely, the existing CER's will also overestimate the variance for the prediction of the cost of a follow-on lot of a particular existing airframe.

Secondly, the Large, et al. study, with its possible loss of information should result in confidence intervals that are larger than necessary. That is, if more information is available when estimating a CER, then the resulting confidence intervals should be tighter.

This study will investigate these hypotheses and thus determine if either of the two existing procedures (LAR and TT) is appropriate or whether a more general procedure is indicated.

Procedure

This thesis will develop a technique to estimate the components of variance model. The technique, based on a method presented by Searle (Ref 12:465-470) is referred to as the RANDOM technique.

If the variance of the error associated with different airframe types is estimated to be zero or near zero, then the TT approach would be appropriate (the correlation is not evident). If the estimated variance of the overall error term is zero and the estimated variance of the error between different airframes is not, then the Large approach is most appropriate. But, if both components of variance are significant, this indicates that neither the LAR nor TT approaches is correct. By implication the RANDOM technique would then be suggested or more appropriate.

This is the task of this study; that is, to develop the RANDOM technique to estimate the components of variance model; to apply the

technique together with the techniques of LAR and TT to the same set of data; and to compare the resulting CER's with the LAR and TT type CER's estimated.

Organisation

This study will first cover the actual development of the RANDOM model estimation technique and its extension to the procedure of generalised least squares. Then, the actual data source and its meaning will be discussed as it is applicable to all three modeling techniques in Chapter III. This study will use a data base consisting of nine fighter/trainer type airframes with a total of 33 lots.

Chapter IV will outline the actual development of all three different sets of CER's. The cost element CER's to be estimated will be that of total engineering hours, total tooling hours, recurring labor hours, and recurring material dollars (1973). This chapter will then present the derived CER's and analyze them.

The next chapter, Chapter V, will address the predictive qualities of all three different types of CER's. All the CER's will predict the respective costs associated with the F-14, lots one, two and at a normalised quantity of 100 airframes. These predictions will then be used for CER comparison. Lastly, in Chapter VI, this work will be summarized and conclusions will be presented.

II. Components of Variance Model

This chapter will first develop the general technique utilized to obtain the CER's for the mixed effects situation. Then it will go into a discussion and short outline of the development of the technique used to estimate the variances. This last portion of the chapter is the crucial part of adapting the components of variance model to airframe cost estimation.

Notation

Matrix notation will be utilized throughout this paper. Capital letters will be used to designate vectors and matrices. The transpose, inverse, and generalized inverse of a matrix will be denoted by A', A^{-1}, and A^{-}, respectively.

An estimator of a parameter, variable, vector or matrix will be denoted by a hat over the symbol. For example, $\hat{A}$ would be an estimate of A.

A vector whose elements have been replaced by the mean of its elements will be indicated by a bar over the symbol, e.g., $\bar{A}$.

The trace of a matrix, which is the sum of the diagonal elements of a matrix, will be denoted by tr, e.g., trA.

Ordinary Least Squares

The standard linear model is given by

$$Y = XB + \epsilon \tag{1}$$

where Y is a vector of observations (dependent variable), X is a vector of explanatory variables, and ϵ is the disturbance vector. The

assumptions made with this model are (Ref 14:110-111):

1. The expected value of the disturbance vector given the values of X is

$$E(\epsilon|X) = 0 \tag{2}$$

2. The variance-covariance matrix of ϵ, given X is

$$VAR(\epsilon|X) = \sigma^2 I \tag{3}$$

where σ^2 is an unknown positive number and I is an n x n identity matrix.

3. The x_i variables (i=1, 1, . . . k) are linearly independent.

4. The n-element disturbance vector, ϵ, is assumed to be normally distributed.

The Ordinary Least Squares (OLS) procedure is used to obtain an estimate of the unknown coefficient vector B so that

$$\hat{Y} = X\hat{B} \tag{4}$$

Now one can define $\hat{\epsilon}$ as a vector of residuals,

$$\hat{\epsilon} = Y - \hat{Y} \tag{5}$$

Without outlining all the well known details the OLS solution for $\hat{B}$ is

$$\hat{B} = (X'X)^{-1} X'Y \tag{6}$$

The derived properties of OLS estimates are (Ref 14:111-115):

1. $\hat{B}$ is unbiased.

2. An unbiased estimator of the covariance matrix for $\hat{B}$ is

$$Var(\hat{B}) = \hat{\sigma}^2 (X'X)^{-1} \tag{7}$$

3. An unbiased estimator for the unknown parameter, σ^2, is given by

$$s^2 = \hat{\sigma}^2 = \frac{\hat{\epsilon}'\hat{\epsilon}}{n-k} \tag{8}$$

Generalized Least Squares

In the standard linear model, Eq (1), it is assumed that the random disturbances are uncorrelated and identically distributed. This results in a diagonal conditional covariance matrix with σ^2 on the diagonal.

Now, if assumption two above is violated, the conditional covariance matrix is no longer diagonal;

$$\text{Var}\ (\epsilon|X) = a^2V \tag{9}$$

where a^2 is an unknown positive parameter and V is a known symmetric positive definite n x n matrix whose trace equals n.

Because V is a symmetric positive definite matrix, its inverse has the same properties. Therefore, there exists a non-singular matrix P such that

$$P'P = V^{-1} \tag{10}$$

To find an estimator of B we transform the standard linear model by multiplying through both sides by P:

$$PY = PXB + P\epsilon \tag{11}$$

The expected value of $P\epsilon$ is still equal to zero and the conditional covariance matrix of $P\epsilon$ given X is:

$$\begin{aligned} \text{Var}\,(P\epsilon|X) &= P\,[\text{Var}\,(\epsilon|X)\,]\,P' \\ &= P\,(a^2V)\,P' \\ &= a^2\,P(P'P)^{-1}\,P' \\ &= a^2 I \end{aligned} \tag{12}$$

So, after the transformation, the new model does meet the assumptions discussed as given under the standard linear model. Therefore, OLS can be utilized to obtain an estimate of the coefficients.

Under GLS an estimate of B is

$$\hat{B} = (X'V^{-1}X)^{-1}\,X'V^{-1}Y \tag{13}$$

with the conditional covariance matrix given by

$$\text{Var}\,(\hat{B}|X) = a^2(X'V^{-1}X)^{-1} \tag{14}$$

and a^2 is estimated by

$$\hat{a}^2 = \frac{1}{n-k}\,(Y-X\hat{B})'V^{-1}(Y-X\hat{B}) \tag{15}$$

where k is the number of explanatory variables used for regression. One can obtain a much more detailed development of GLS estimators from Theil (Ref 14:237-240).

Extension of GLS to Airframe Cost Estimation

As stated in the introduction to this paper, if one assumes that there are two sources of error in airframe cost estimation, then somehow the CER's developed must take this into account. These two error terms are assumed to be normally distributed and independent of one another. This essentially causes two problems. (1) The conditional covariance matrix is not diagonal and therefore GLS must be used to obtain the CER's. (2) One must obtain an estimate of the two different

variances to obtain an estimate of V. Once an estimate of V is obtained, then it can be used to replace V in Eq (13) to estimate the elements of B.

Now, let us assume that the components of variance model to be used is

$$y_{ij} = X_{ij} B + u_j + \epsilon_{ij} \tag{16}$$

u_j is the error term associated with different type airframes and ϵ_{ij} is the overall error term.

It is appropriate here that we discuss the relevance of two error terms in this new model. The error associated with different aircraft airframes could be tied to several factors. The error could result from totally new aircraft designs, new materials used, and new and more complicated design features, etc. These aircraft differences would definitely affect cost but they may not be totally captured by the explanatory variables utilized in the CER's. The overall error term, on the other hand, would most probably reflect the general economic conditions present under different contracts, e.g., small changes in labor costs caused by inflation or strikes, slight changes in design from one lot to another, etc.

Now it will be assumed that the distributions of both u_j and ϵ_{ij} are normal, independent of one another, and have variances equal to σ_U^2 and σ_ϵ^2, respectively (Ref 16 :13).

$$u_j \sim N(\alpha, \sigma_U^2)$$

$$\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$$

$$E\left[(u_j - \alpha)\, \epsilon_{ij}\right] = 0 \text{ for all } i, j$$

Now, let g represent the overall error term for the model:

$$g_{ij} = u_j + \epsilon_{ij} \tag{17}$$

where g_{ij} is distributed normally with a mean, α, and a variance equal to $\sigma_U^2 + \sigma_\epsilon^2$:

$$g_{ij} \sim N(\alpha, \sigma_U^2 + \sigma_\epsilon^2)$$

Then the variance of g_{ij} is equal to

$$\text{Var}(g_{ij}) = \begin{bmatrix} V_1 & & & & 0 \\ & V_2 & & & \\ & & V_3 & & \\ & & & \ddots & \\ 0 & & & & V_r \end{bmatrix}$$

where r equals the number of different airframes, zeroes are on the off main diagonal, and V_j is a symmetric matrix.

$$V_j = \begin{bmatrix} \sigma_U^2 + \sigma_\epsilon^2 & \sigma_U^2 & \cdots & \sigma_U^2 \\ \sigma_U^2 & \sigma_U^2 + \sigma_\epsilon^2 & \cdots & \vdots \\ \vdots & \sigma_U^2 & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_U^2 & \cdots & \sigma_U^2 & \sigma_U^2 + \sigma_\epsilon^2 \end{bmatrix}$$

where $j = 1, 2, \ldots, r$ and the size of the matrix is m x m with m equal to the number of lots produced of the jth aircraft.

The development of the value for the off-diagonal elements of V_j follows easily from the basic assumptions concerning u_j and ϵ_{ij}. The off-diagonal elements correspond to the covariance between g_{ij} and g_{k_j}, i.e., the covariance between two different lots of the same type aircraft.

$$\text{Covariance } (g_{ij}, g_{k_j})_{k \neq i}$$

$$= E\left[(u_j - \alpha + \epsilon_{ij})\ (u_j - \alpha + \epsilon_{k_j})\right]$$

$$= E\left[(u_j - \alpha)\ (u_j - \alpha) + (u_j - \alpha)\ \epsilon_{ij} + (u_j - \alpha)\ \epsilon_{k_j} + \epsilon_{ij}\ \epsilon_{k_j}\right] \quad (18)$$

Since the assumptions cause $E\left[(u_j - \alpha)\ \epsilon_{ij}\right]$, $E\left[(u_j - \alpha)\ \epsilon_{k_j}\right]$, and $E\left[\epsilon_{ij}\ \epsilon_{k_j}\right]$ to equal zero, then the covariance $(g_{ij}, g_{k_j}) = E\left[(u_j - \alpha)(u_j - \alpha)\right]$, which by definition equals σ_U^2.

The general model therefore becomes

$$Y = XB + G \quad (19)$$

with G distributed normally with a mean, α, and a variance equal to V:

$$G \sim N(\alpha, V)$$

If one then has estimates for the two variance components, then it would be a simple matter to build the $\hat{V}$ matrix and then estimate the coefficients of the model (CER) using Eq (13) (Ref 16:13-14).

Estimation of the Variance Components

The actual thrust of this thesis is built around the variance components. Up to this point nothing has been said about the actual

method utilized to estimate these values. After an extensive literature search (see Supplementary Bibliography, Ref 1, 2, 3, 4, 5, 7, 8), a method reported by Searle was chosen (Ref 12:465-470). This method of variance component estimation uses the fitting constants method (commonly referred to as Henderson's Method 3) and an iterative technique developed by Thompson involving the ratio $\sigma_\epsilon^2/\sigma_U^2$ (Ref 15:767-773). The estimated components obtained from the technique are both unbiased and maximum likelihood. This technique was the only one discovered that obtains estimates with the above properties when the data is unbalanced.

Outline of Variance Component Estimation Technique

The following portion of this chapter is a very general development of the variance component estimation technique. The general model is

$$Y = XB + ZU + \epsilon \tag{20}$$

where in our case, the vector Y is natural logarithm (ln) of cost (dependent variable), X is the ln of the matrix of explanatory variables, which is of rank r, Z is a matrix of indicator variables (ones and zeroes) of full rank, B is a vector of unknown fixed constants, U represents the effects of a single random factor (in our case different type aircraft) having a variance of σ_U^2, and ϵ is the random disturbance term over all aircraft and lots (Ref 12:465).

For airframe cost estimation the X and Z matrices will be constructed as follows:

$$X = \begin{bmatrix} \ln S & \ln W & \ln Q & \cdots \\ \downarrow & \downarrow & \downarrow & \end{bmatrix}$$

where the number of rows is equal the number of separate airframe lot observations. One will notice that a column of ones was not used. The Z matrix will consist of as many columns as there are different airframes and in each column there will be ones, where the number of ones will correspond to the total number of lots of that particular airframe built. The ones will be in the rows corresponding to the rows of the respective costs. The Z matrix columns will sum to one, e.g.,

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

Now, the estimators for σ_ϵ^2 and σ_U^2 are

$$\hat{\sigma}_\epsilon^2 = \frac{Y'Y - [R(U) + R(B|U)]}{n - r(X)} \tag{21}$$

and

$$\hat{\sigma}_U^2 = \frac{R(U) + R(B|U) - R(B)}{\text{tr}\left[Z'Z - Z'X\,(X'X)^{-}X'Z\right]} \tag{22}$$

where n is the total number of observations and the R's denote the sum of squares of reduction due to whatever symbol is contained in the parentheses. These results provide an iterative procedure because the reductions R(U) and R(B|U) involve $\lambda = \hat{\sigma}_\epsilon^2/\hat{\sigma}_U^2$. The estimation is done by picking an initial value of λ (zero is acceptable) and then calculating Eqs (21) and (22); recalculating λ and so on until convergence. A detailed development of this procedure and the resulting equations for the different reductions are contained in Appendix A (Ref 12:465-470).

The combination of the components of variance model Eq (16), the generalized least squares technique, and Searle's method of estimating the variance components will be referred to as the RANDOM technique throughout the remainder of this study.

Relationship of the RANDOM Technique to the Techniques of Timson and Tihansky and Large, et al.

This thesis is comparing the results of three different CER development techniques. But, actually the RANDOM technique alone could indicate the better of the three methods. The variance components estimated as outlined above should give a very good indication of which method is more appropriate for a particular CER.

One can see from the construction of the V_j matrices and the V matrix (see pages 13-14) that if $\hat{\sigma}_U^2$ is equal to zero, the V matrix reduces to the diagonal form assumed in Timson and Tihansky's (TT's) study. This would lead one to believe that the TT CER's are best.

On the other hand, if $\hat{\sigma}_\epsilon^2$ is equal to zero, then the V_j matrices would be constructed with the value, $\hat{\sigma}_U^2$, only. This, of course, would make all the V_j's singular and this in turn would lead to a singular V matrix. The solution to this singularity problem would then be to limit each V_j matrix to a one by one matrix with a value of $\hat{\sigma}_U^2$ which reduces the V matrix to a square matrix with $\hat{\sigma}_U^2$ on the diagonal and seroes elsewhere. The row/column size of this matrix would be equal to the number of different type airframes. In other words, with $\hat{\sigma}_\epsilon^2 = 0$ one is now faced with the Large, et al. (LAR) approach.

Now, as an extension of the above relationships, if there is no great difference in magnitude between $\hat{\sigma}_U^2$ and $\hat{\sigma}_\epsilon^2$, then one could deduce

that neither the LAR nor TT techniques is best. In this case one could argue, then, that somewhere in between LAR and TT is the better technique, RANDOM.

A Possible Test of Hypothesis

One could carry the theory put forth in the previous section one step further. If a valid test statistic were available, one could use a statistical test of hypothesis to determine if one or the other of the components is equal to zero with some specified probability. For example, the null hypothesis could be

$$Ho : \sigma_U^2 = 0$$

or

$$Ho : \sigma_\epsilon^2 = 0$$

against the respective alternate hypothesis

$$Ha : \sigma_U^2 \neq 0$$

or

$$Ha : \sigma_\epsilon^2 \neq 0$$

Unfortunately, at present, there is no test statistic derived which is directly applicable to the variance components to be estimated in this study.

One could make an assertion here, though, that if the estimated components are of near equal magnitude, there is no strong evidence that either component is equal to zero and the other is not.

III. The CER's to be Developed and the Data

Four airframe cost elements will be addressed in this paper. These are engineering hours (total), tooling hours (total), labor hours (recurring), and material dollars (recurring). There are two reasons these four elements were chosen over any others. First, the data for these cost elements is much more readily available and more consistent than the data for the other elements. Secondly, the CER's estimated for these four cost elements will be adequate to determine the usefulness and applicability of the RANDOM technique as compared to the other two. If the RANDOM technique proves to be good for this set of CER's, then this will be sufficient cause for investigation of the technique into all areas of airframe cost estimation.

The first portion of this chapter will discuss the functional form of the CER's and the independent (explanatory) variables to be utilized in this study. The last sections of this chapter will focus on the source and reduction of the actual data to be utilized.

Physical Characteristics (Explanatory Variables) Related to Cost

The selection and acquisition of military aircraft begins with little more than a requirement for a certain performance and/or a specific set of physical attributes (maximum speed, weight, etc.). Prior to 1975 it was generally thought that, of the characteristics available in the early conceptual stages of aircraft development, the only ones with any statistically proven parametric significance were airframe unit

weight, maximum speed at best altitude, production rate and production quantity. (Airframe unit weight is defined as empty weight minus the following: wheels, brakes, tires, and tubes; engines--main and auxiliary; rubber or nylon fuel cells; starters--main and auxiliary; propellers; auxiliary power-plant unit; instruments; batteries and electrical power supply and conversion; avionics group; turrets and power-operated mounts; air conditioning, anti-icing and pressurization units and fluids; cameras and optical view finders; trapped fuel and oil.) (Ref 8:19). Therefore, the CER's developed most often use different combinations and subsets of these four explanatory variables.

Functional CER Forms

The functional forms for the cost estimating equations themselves that conform best to airframe cost data and any least squares technique are the log-linear and exponential forms:

$$\text{(Log-linear) } Y = e^{a}S^{b}W^{c}e^{\epsilon} \tag{23}$$

$$\text{(Exponential) } Y = e^{a}S^{b}W^{c} + \epsilon \tag{24}$$

where Y is the dependent variable (cost to be estimated), W denotes unit weight, S denotes maximum aircraft speed at best altitude, e is the base of the natural logarithms, ϵ is the normally distributed error term, and a, b, and c are the constants to be estimated. The logarithmic form can be "linearized" for the application of the least squares method by taking the natural logarithm of both sides of the equations:

$$\text{(Log-linear) } \ln Y = a + b \ln S + c \ln W + \ln \epsilon \tag{25}$$

The only difference between the two forms is the manner in which the error term ϵ is treated (multiplied or added).

Of the two forms listed in the preceding paragraph, the log-linear is most often used (Ref 8; Ref 13). This is mainly due to the fact that this form conforms better to the real world. The error distribution is skewed so that at a given confidence level (a statistically determined interval within which a CER estimate of cost is said to lie between with a certain probability) the lower bound of the interval is not as great as the upper bound which is analogous to the chances being greater for a cost overrun on a contract than an underrun (for the linear or exponential models, the maximum underrun and overrun are of equal magnitude). Also, the range of the logarithmic CER is from zero to positive infinity (no negative cost) whereas the exponential CER ranges from minus infinity to plus infinity. There are other lesser advantages to the log-linear form and they are explained in the report by Timson and Tihansky (TT) (Ref 13:4-13).

This study will use the same three variables for all CER's estimated:

1. Maximum speed at best altitude (S)
2. Unit airframe weight (W)
3. Quantity of airframes produced (Q).

The Large, et al. (LAR) CER's do not, of course, use Q explicitly. Therefore, a version of the LAR approach, which does use Q, is developed for comparison. These variables are representative of the variables used for most previously developed CER's in this area of study. Since the same variables will be used for all three types of CER's, the comparison of the approaches should be valid.

Data Source

The data base used by this report was compiled by the RAND Corporation in conjunction with the LAR study done for the Assistant Secretary of Defense (Ref 8). The data is contained on work sheets prepared by RAND and these are kept on file at the Cost Library, Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio.

Data Reduction

The data actually utilized to develop all the CER's is a small subset of the total base available from RAND. Nine airframe types are included and the total number of lots is equal to 33. The aircraft included in this subset are fighter and trainer types only. These aircraft and the respective number of lots are:

A-4D (3)	F-105B,D (4)	F-4A,B (4)
F-102A (4)	F-106A,B (5)	T-38A (3)
F-104A,B,C (3)	A-5A,C (4)	F-111A,C,E (3)

The cost data for these airframes is presented in Appendix C. The number of airframes and the cost of each lot is adjusted for major design changes. Also, where the number of airframes in actual contracts is very small and no major design changes were made, these small contracts were combined with preceding or succeeding contracts to provide a more consistent observation set for use in the CER development. The LAR and TT studies both proceeded in this same manner.

Handel, in a thesis on airframe cost estimation utilized the same procedures as above and his data set included all nine of the airframes to be used in this thesis. After careful examination of his data and

referencing the RAND work sheets, it was determined that his data would be adequate for this study (Ref 4:26-34).

The data for the four cost elements to be studied can be broken down into recurring and non-recurring dollars/hours. Previous studies by the RAND Corporation have indicated that, where possible, recurring cost data should be utilized for CER development. But, in these same studies discrepancies were discovered in the contractor reported data for tooling and engineering costs. As a result, only total tooling and engineering costs/hours were estimated in this study (Ref 8:18). The Labor and Material hours/costs are of the recurring type only.

Cost Element Definitions

The tooling costs consist of the material, labor, and overhead costs for the assembly tools, dies, jigs, fixtures, work stands, and test equipment needed for the production of an airframe. The number of direct labor hours used for tooling is highly correlated with the tooling cost and therefore CER's can be developed using tooling hours. The tooling cost, then, can be obtained by multiplying the number of direct tooling hours by a composite hourly rate, taking into account all those cost items mentioned earlier (Ref 4:26-27).

The total engineering costs consist of the preliminary design effort and integration and of the material, labor, and overhead costs expended in the engineering for the basic airframe and of the system engineering performed by the prime contractor. Total engineering hours can be utilized in the exact same way as tooling hours, with the total engineering cost estimated by applying a composite rate to the number of engineering labor hours expended (Ref 4:27-28).

The recurring cost of manufacturing material includes the costs of raw and semifabricated materials, purchased parts and purchased equipment. The dollar cost of the manufacturing material was adjusted to 1973 dollars with the use of the price adjustment indices listed in Table I, so any material recurring costs estimated by CER's in this study are the constant year dollar sum of material, purchased equipment, and government furnished equipment (Ref 4:32).

Lastly, manufacturing recurring labor hours consist of the direct labor required to machine, process, assemble, and fabricate the major structure of an airframe (Ref 4:29).

The CER's

In this thesis there will be actually six different CER's estimated for each cost element. These CER's differ with respect to the technique used (RANDOM, LAR, or TT) and/or the way in which the data is utilized (data/cost approach).

Cumulative Cost Approach. Cumulative airframe cost and cumulative airframe quantity lot data will be used with both the RANDOM and TT techniques to develop a cumulative RANDOM CER and a cumulative TT CER. The speed and weight used will be that of the last airframe produced in each lot.

Marginal Cost Approach. The second cost approach will again be utilized by both the RANDOM and TT techniques. This approach will use average cost per airframe per lot. To obtain this average cost per lot the actual lot cost is divided by the number of airframes produced in the lot. This can be derived from the cumulative cost data presented in Appendix C. This cost will be termed Marginal Cost in this thesis.

Table I
Price Adjustment Indices
1973

Year	Material	Equipment
1952	2.625	2.972
1953	2.480	2.808
1954	2.359	2.656
1955	2.224	2.506
1956	2.081	2.353
1957	1.970	2.226
1958	1.859	1.078
1959	1.793	1.981
1960	1.718	1.892
1961	1.672	1.833
1962	1.614	1.756
1963	1.579	1.696
1964	1.528	1.632
1965	1.479	1.568
1966	1.422	1.496
1967	1.359	1.422
1968	1.295	1.343
1969	1.208	1.249
1970	1.177	1.188
1971	1.137	1.138
1972	1.094	1.081
1973	1.000	1.000

This, of course, results in two different CER's also - the Marginal RANDOM and Marginal TT CER's.

There is a good reason for deriving these marginal CER's. This study has acknowledged the fact that there may be correlation in the data from the lots of the same type airframe. The RANDOM technique will hopefully account for this fact. But, there may be another type of correlation taking place when one uses the cumulative cost approach. Auto-correlation could result from the fact that each successive lot cost (cumulative) includes the cost of the preceding lot(s). If present, this auto-correlation affects the statistics associated with the cumulative CER's. One common result of this type of data correlation is a larger than actual coefficient of determination, R^2, thus indicating that the CER fits the data better than it actually does. The marginal cost approach eliminates the possibility of this auto-correlation.

The transformation of the RAND explanatory variable data to marginal data is quite involved though. The variables of weight and speed will refer to the same values as mentioned in the cumulative cost approach definition, but the airframe quantity explanatory variable will refer to a quite different value. In this marginal cost approach Q will symbolize the true lot mid-point of each lot. Again, this approach will only be applied with the RANDOM and Timson and Tihansky techniques.

To discuss the meaning and derivation of true lot midpoints one must discuss the existence of a production learning curve. In the aerospace industry, for example, there is empirical evidence that there exists a learning process or phenomenon that causes a reduction in production cost as the number of items produced increases. Although there

are several hypotheses on the exact manner in which this reduction takes place, the basis of learning-curve theory is that each time the total quantity of items produced doubles, the cost per item is reduced to some constant percentage of the previous cost (Ref 1:1).

Most learning curve slope derivations deal with cumulative average cost:

$$y_c = ax^b \tag{26}$$

where a is the cost of the first item produced, x is the cumulative number of items produced, b is an exponent that measures slope, and y_c is the average cost of all items produced up to and including x. The learning-curve slope, s, describes the average cost of 2x items as a fraction of the average cost of x items and is related to b as follows:

$$s = 2^b \quad \text{or} \quad b = \frac{\text{Log } s}{\text{Log } 2} \tag{27}$$

Now, in our case we are dealing with total cumulative cost and this changes Eq (26):

$$t = y_c x = \text{total cost of x items} \tag{28}$$

and therefore

$$\begin{aligned} y_c &= ax^b x \\ &= ax^{b+1} \end{aligned} \tag{29}$$

The above changes Eq (27) as follows:

$$b = \frac{\text{Log } s}{\text{Log } 2} + 1 \tag{30}$$

or

$$b - 1 = \frac{\text{Log } s}{\text{Log } 2} \tag{31}$$

With the above results then one can estimate the unit learning-curve slope given some value for b by using the equation

$$s = e^{(b-1)\,\text{Log } 2} \qquad (32)$$

This is the equation used by this study to estimate the slope of the unit learning curve for all four cost elements. The value used for b in this report is the exponent of the airframe quantity explanatory variable as estimated by the RANDOM technique under the cumulative cost approach.

Once an estimate of the unit learning-curve slope is obtained using Eq (32) then the learning-curve tables derived by Boren and Campbell of the RAND Corporation can be utilized to obtain the true lot midpoints. The procedure and theory are given in much greater detail in the publication titled Military Equipment Cost Analysis, and also in the Boren and Campbell volumes of learning-curve tables (Ref 11:93-125; Ref 1: Vol. 1, 1-13).

One can see from the above explanation that this report's estimate of s involves an estimate of b. To be sure, there is a great possibility of significant error. But, because we are dealing with relatively small lot quantities, even a large estimation error results in true lot midpoint errors of only one or two airframes.

LAR Technique/Cost Approach. The fifth and sixth CER's to be presented can be described as being derived with either different techniques and/or different cost approaches. The fifth CER derivation will be by the use of the LAR technique as presented in their report. This technique can best be described by the way in which data is utilized (cost approach). The cost will be the cumulative cost (estimated) of a quantity of 100 airframes. This will be called the LAR CER.

The historical airframe cost data very seldom has an actual lot cost observation at cumulative quantity 100, so the following procedure

was utilized in the LAR and Handel studies to obtain the observations. They assumed a log-linear relationship for learning. Then actual cumulative costs versus the corresponding cumulative quantities were plotted on logarithmic graph paper for each airframe type. A straight line was then drawn through these plot points immediately preceding and succeeding quantity 100. The cost observation for this quantity was then read off the graph (Ref 8:45). This method implies a cumulative average learning curve as compared to the unit learning curve assumed in the marginal approach explained earlier. Even with these two different assumptions, the LAR CER's and the two marginal CER's may be compared because the differences would be very slight in the lot sizes used in this study.

The last CER to be developed in this study will be a cumulative total cost CER. Now the technique/cost approach used to derive this relationship is very similar to the technique/cost approach described for the LAR CER. The difference is that total airframe program cost will be used as the dependent variable and the total airframe quantity will be explicitly utilized as an explanatory variable, e.g., the total cost and quantity of all the F-4 airframes produced. This CER is developed to compare the learning curve approach of LAR's technique to a very similar technique, i.e., the same number of observations, but no estimation of a learning curve to obtain some normalized airframe quantity data for regression.

Summary. The preceding sections define the six CER's to be developed and analyzed in this study. Again, they are the:

1. Cumulative RANDOM CER
2. Cumulative TT CER
3. Marginal RANDOM CER

4. Marginal TT CER

5. LAR CER

6. Cumulative total cost CER

One must remember that the LAR and cumulative total cost CER techniques actually involve different cost approaches (they both utilize the ordinary least squares statistical technique). The TT and RANDOM techniques utilize airframe lot observations and therefore cannot be utilized with the LAR or cumulative total cost data approaches. Conversely, the LAR and cumulative total cost techniques cannot use the cumulative or marginal lot data.

Tables II and III list all the airframe types and the associated data to be used to estimate all six CER's.

Table II
Aircraft Characteristics (Explanatory Variables)

Aircraft/Lot #	Cumulative Approach			Marginal Approach			Large, et al. Approach Quantity = 100		Total Program Cost and Quantity Approach		
	S	W	Q	S	W	Q*	S	W	S	W	Q
A-4/1	578	4971	20	578	4971	7					
/2	578	5102	72	578	5102	42	578	5072	578	5072	166
/3	578	5072	166	578	5072	114					
F-102/1	677	12000	79	677	12000	20					
/2	677	12052	99	677	12052	90					
/3	677	12052	207	677	12052	148	677	12304	677	12052	235
/4	677	12052	235	677	12052	222					
F-104/1	1150	8000	25	1150	8000	8					
/2	1150	8069	234	1150	8069	99	1150	7963	1150	8170	255
/3	1150	8170	255	1150	8170	245					
F-105/1	1195	19733	80	1195	19733	20					
/2	1195	18911	108	1195	18911	94					
/3	1195	17954	176	1195	17954	141	1195	19301	1195	17960	315
/4	1195	17960	315	1195	17960	241					
F-106/1	1153	14630	96	1153	14630	23					
/2	1153	15400	130	1153	15400	113					
/3	1153	14784	170	1153	14784	150	1153	14620	1153	15105	323
/4	1153	15269	175	1153	15269	173					
/5	1153	15105	323	1153	15105	243					

Table II (continued)

Aircraft/Lot #	Cumulative Approach			Marginal Approach			Large, et al. Approach Quantity = 100		Total Program Cost and Quantity Approach		
	S	W	Q	S	W	Q*	S	W	S	W	Q
A-5/1	1147	22222	25	1147	22222	8					
/2	1147	23828	77	1147	23828	48	1147	23499	1147	26613	120
/3	1147	26459	97	1147	26459	88					
/4	1147	26613	120	1147	26613	109					
F-4/1	1220	17320	47	1220	17320	13					
/2	1220	18066	119	1220	18066	79	1220	17220	1220	18342	338
/3	1220	18244	191	1220	18244	154					
/4	1220	18342	338	1220	18342	259					
T-38/1	750	5362	20	750	5362	7					
/2	750	5348	70	750	5348	42	750	5376	750	5482	214
/3	750	5482	214	750	5482	131					
F-111/1	1262	32926	159	1262	32926	36					
/2	1262	33000	183	1262	33000	172	1262	33150	1262	33000	277
/3	1262	33000	277	1262	33000	228					

*The Q variable has different values associated with each cost element (the learning curve slope changes). The values given above are for engineering only. See Table III for other cost element Q values.

Table III
Remaining Quantity Values for the Marginal Cost Approach

Aircraft	Labor Q	Tooling Q	Material Q
A-4	8	7	8
	39	42	44
	116	114	116
F-102	25	21	28
	90	89	90
	149	148	150
	222	222	222
F-104	9	8	10
	106	101	110
	245	245	245
F-105	26	21	28
	95	94	95
	141	141	141
	242	241	242
F-106	30	25	33
	113	113	113
	150	150	150
	173	173	173
	245	244	245
A-5	9	8	10
	49	48	49
	88	82	88
	109	109	109
F-4	16	13	17
	80	79	81
	154	154	154
	261	260	261
T-38	8	7	8
	43	42	43
	134	132	135
F-111	49	39	54
	172	172	172
	229	228	229

IV. CER Development, Presentation, and Analyses

To reiterate, this thesis focuses on the RANDOM technique of developing airframe CER's. To analyze the results of this development there must be some frames of reference. The frames utilized in this study are two past airframe cost estimating reports, the one by Timson and Tihansky (TT) and the other by Large, Campbell, and Cates (LAR). To assure the proper comparisons all three techniques were used to develop the exact same CER's using the exact same data. The comparisons, then, that will follow in this chapter will be as equitable and valid as is possible.

CER Values to be Presented

The actual presentation of the CER's will include all the coefficients estimated and several associated statistics. The statistics will be the coefficient of determination, R^2, for each CER, the Student t-ratio for each coefficient, the estimated residual variance, $\hat{\sigma}_R^2$, for the TT and LAR CER's, and finally both $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_U^2$ for the RANDOM type CER's (the variance components).

The coefficient of determination is defined as:

$$R^2 = 1 - \frac{(Y - \hat{Y})' (Y - \hat{Y})}{(Y - \bar{Y})' (Y - \bar{Y})} \tag{33}$$

This particular R^2 statistic is uncorrected for degrees of freedom but it will serve a useful purpose for CER comparisons. This will tend to

yield relatively higher R^2 values for the LAR CER's since they are characterized by few degrees of freedom.

The Student t-ratio is used to test the hypothesis that a coefficient equals zero or otherwise. The t-ratios will be placed in parentheses below the respective coefficients in the tabular presentation of the CER's. The t-ratio presented here is a standard statistic associated with each estimated CER coefficient. It is calculated by dividing the sum of squares due to regression (for a particular coefficient) by the residual standard deviation, $\hat{\sigma}_R$. It is the test statistic for determining the significance level of each estimated coefficient. Basically, the larger the statistic, the greater the significance.

The variance components, $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_U^2$, were estimated using the Searle technique and all the estimations were obtained after 10 iterations. An example computer program is provided in Appendix B. The determination of 10 iterations was made after an extensive number of computer runs. In all cases except the Engineering Marginal cost approach the variance component values converged rapidly and after five iterations the values changed beyond the fifth decimal place only. The number 10 was chosen for even more accuracy (Ref 2 :23-25).

The symbols, variables, and/or acronyms used for the tabular presentation of the CER's (excluding some which are already defined) are presented in Table IV and explained in the succeeding paragraph.

The variables and/or acronyms defined in Table IV will be used as follows. The name of a particular CER will be denoted by a combination of symbols, always beginning with the symbol for the applicable cost element, then followed by the symbol describing the cost approach,

Table IV
Definitions

Acronym and/or Variable	Definition
E	Total Engineering Hours
T	Total Tooling Hours
L	Recurring Manufacturing Labor Hours
M	Recurring Material Dollar Cost (1973 dollars)
R	RANDOM Technique
T	Timson and Tihansky Technique
L	Large, Campbell, and Cates Technique
C	Cumulative Cost Approach
M	Marginal Cost Approach
CUMT	Cumulative Total Program Cost Approach
K	Constant Coefficient
Q	Quantity of Airframes (as explained earlier)
S	Maximum Speed at Best Altitude (Knots)
W	Unit Weight (lbs.)

then finally followed by the symbol denoting the technique used to derive the particular CER, e.g., LCR will denote the total Labor hours CER using the Cumulative cost approach and RANDOM technique for its derivation. Two other examples would be MMT which means Material, Marginal, Timson and Tihansky, and LCUMT which denotes Labor hours using the CUMulative Total program cost approach/technique.

Total Engineering Hours

Thirty-three observations of nine fighter type aircraft were used to derive these CER's. The estimation of the RANDOM technique variance components utilized 10 iterations. The resulting CER's are presented in Table V.

The t-ratio statistics for all the engineering hour CER's except the cumulative total CER, indicate that of the four explanatory variables used weight and quantity are by far the most significant.

The cumulative RANDOM CER variance component, $\hat{\sigma}_U^2$, is much larger than $\hat{\sigma}_\epsilon^2$. This indicates that under the cumulative cost approach the LAR technique is appropriate. This is consistent with the assumption that there is between lot correlation due to both random effects and auto-correlation. The LAR technique implicitly takes these factors into account and therefore may be the technique best suited to engineering cost estimation.

All the cumulative CER's show inflated R^2 values which further indicates that auto-correlation is present when one uses the cumulative cost approach.

The R^2 associated with the cumulative total CER is larger than the LAR R^2. This is understandable since the cumulative total approach does

Table V
CER Coefficients and CER Statistics
Engineering Hours

CER Type	Estimated Coefficients for the Explanatory Variables K	S	W	Q	Statistics R^2	$\hat{\sigma}_R^2$	$\hat{\sigma}_\epsilon^2$	$\hat{\sigma}_U^2$
ECR	-.7931 (-.55)	.7859 (2.56)	1.0412 (7.54)	.2476 (12.35)	.999	-	.0059	.0279
ECT	-.2429 (-.31)	.5209 (3.11)	1.1735 (14.68)	.2495 (6.15)	.966	.0302	-	-
ECL	3.7937 (1.67)	.5975 (1.22)	.8350 (3.70)	-	.893	.0757	-	-
EMR	-2.8788 (.97)	1.0905 (1.69)	1.0024 (3.29)	-.9732 (-10.29)	.656	-	.3450	.0279
EMT	-2.7191 (-1.0)	1.0519 (1.8)	1.0164 (3.63)	-.9809 (-10.17)	.816	.3687	-	-
ECUMT	-1.0757 (-.61)	.7578 (1.82)	1.0847 (6.26)	.2566 (1.06)	.970	.0416	-	-

not estimate observations as the LAR approach does. This cumulative total technique also accounts for the correlation problems in the same manner as the LAR technique. These findings, then, indicate that the cumulative total CER is even more appropriate than the LAR CER for an engineering hour cumulative CER.

The marginal RANDOM CER variance components indicate the reverse of the respective cumulative findings. That is, $\hat{\sigma}_{\epsilon}^2$ is much larger than $\hat{\sigma}_{U}^2$. This indicates that under the marginal approach for engineering, the TT technique is most appropriate. This finding is not consistent with the hypotheses put forth in this study. This would indicate that there is no data correlation problem.

The R^2 values for the marginal CER's are more believable (smaller than the cumulative R^2's). This is a result of the fact that autocorrelation is no longer present here.

This set of CER's did show significant inconsistencies as compared with the others. The iterative routine for estimating the variance components for the marginal RANDOM CER did not converge well. This was the only CER development that displayed this characteristic. This could account for the marginal result indicating the TT technique is appropriate when in fact it should not be. The data is the cause of these problems. One can see that the non-recurring costs are heavily loaded into the first lot of all the airframe types; much more so than in the other cost categories. Recall, the engineering costs are total costs because data inconsistencies prevented the separation of recurring and non-recurring elements (see page 24).

Because of the problems with the engineering marginal data mentioned above and correlation, the most appropriate engineering CER indicated here is the cumulative total CER.

Total Tooling Hours

Again, here, 33 observations were utilized for the development of all the tooling hour CER's excluding the LAR and cumulative total types for which nine observations were used. Ten iterations were used to estimate the variance components. Table VI presents the results.

The most significant coefficients for these CER's are the constant coefficient, airframe unit weight, and quantity (where applicable).

The estimated variance components for the cumulative RANDOM CER again show that the greatest source of error is due to different types of airframes in the observation set, i.e., $\hat{\sigma}_U^2$ is much larger than $\hat{\sigma}_\epsilon^2$. As stated for the engineering CER's this indicates that the LAR CER is more appropriate when one uses cumulative data.

The cumulative R^2's again show inflation due to auto-correlation. The R^2 for the cumulative total cost CER is greater than the LAR CER R^2, where the observations are estimated. Therefore, under the cumulative data approach the cumulative total cost CER is better than the RANDOM, LAR and TT CER's.

The RANDOM variance components estimated under the marginal data approach are very nearly equal in magnitude. This is consistent with previous assumptions made. The marginal approach accounts for auto-correlation and the RANDOM technique uses the most information (data) while explicitly accounting for the between-lot random effects. These factors indicate that neither the LAR nor TT techniques are appropriate,

Table VI
CER Coefficients and CER Statistics
Tooling Hours

CER Type	Estimated Coefficients for the Explanatory Variables				Statistics			
	K	S	W	Q	R^2	$\hat{\sigma}_R^2$	$\hat{\sigma}_\epsilon^2$	$\hat{\sigma}_U^2$
TCR	9.6203 (2.4)	-.1040 (-.13)	.5669 (1.7)	.3135 (11.6)	.996	-	.0122	.2694
TCT	10.5183 (10.0)	-.6206 (-2.73)	.7914 (7.3)	.4265 (7.76)	.858	.0554	-	-
TCL	12.7794 (6.73)	-.7923 (-1.94)	.8914 (4.74)	-	.825	.0527	-	-
TMR	11.0981 (2.34)	-1.0051 (-.98)	1.0418 (2.21)	-.8974 (-11.71)	.887	-	.2394	.3040
TMT	11.5264 (4.82)	-.9925 (-1.93)	.9598 (3.89)	-.8347 (-9.77)	.777	.2857	-	-
TCUMT	10.0921 (5.89)	-1.1431 (-2.81)	.9538 (5.62)	.8591 (3.63)	.910	.0399	-	-

as hypothesized. The more general RANDOM technique is best under the tooling cost category.

The marginal R^2 values are again more believable than the respective cumulative R^2's. Also, the RANDOM R^2 is the largest which further defends the marginal RANDOM approach.

In light of the above results for the tooling cost category, the marginal RANDOM CER is recommended.

Recurring Labor Hours

The same type airframe data is used for the labor CER's as was used for the previous two sets. The results are shown in Table VII. Ten iterations were used for the estimation of the variance components.

The t-ratio statistics point to the same coefficients as most significant here as in the tooling CER's.

The cumulative RANDOM CER statistics indicate again that the LAR technique is most appropriate for cumulative data. The value $\hat{\sigma}_U^2$ is greater by more than a factor of ten than $\hat{\sigma}_\epsilon^2$. The cumulative R^2 values continue to show inflation due to auto-correlation.

As noted in the previous two cost category presentations, the cumulative total CER shows a better fit than the LAR CER which leads to the same conclusion as before. That is, if one uses cumulative data, then the cumulative total cost technique is the most appropriate.

The RANDOM CER estimated with the marginal data does not indicate the superiority of either the LAR or TT techniques, since the components of variance are nearly equal.

Table VII
CER Coefficients and CER Statistics
Labor Hours

CER Type	Estimated Coefficients for the Explanatory Variables				Statistics			
	K	S	W	Q	R^2	$\hat{\sigma}_R^2$	$\hat{\sigma}_\epsilon^2$	$\hat{\sigma}_U^2$
LCR	4.1595 (1.69)	-.1705 (-.34)	1.1093 (5.27)	.6104 (33.7)	.999	-	.0050	.0927
LCT	3.8358 (3.5)	-.1400 (-.59)	1.1092 (9.81)	.6357 (11.08)	.938	.0603	-	-
LCL	7.0597 (2.78)	-.0856 (-.16)	1.0373 (4.12)	-	.859	.0944	-	-
LMR	4.3113 (2.2)	-.6165 (-1.45)	1.3781 (7.03)	-.3958 (-10.39)	.988	-	.0516	.0477
LMT	4.1231 (3.59)	-.5281 (-2.13)	1.3280 (11.19)	-.3820 (-8.79)	.887	.0660	-	-
LCUMT	3.6022 (1.81)	-.5639 (-1.19)	1.2600 (6.4)	.9459 (3.45)	.951	.0538	-	-

Thus, for the labor cost element the marginal RANDOM technique provides the best CER estimate considering the data utilized in this study.

Recurring Material Dollars

The results in this section very closely parallel the results for the tooling and labor CER's presented in the previous two sections. The same data was used. The most significant coefficients in all the CER's are weight and quantity. The cumulative RANDOM statistics and the R^2 values again point to the cumulative total CER as the most appropriate cumulative type CER. The marginal RANDOM statistics, $\hat{\sigma}_U^2$ and $\hat{\sigma}_\epsilon^2$ are of near equal magnitude and therefore indicate the marginal RANDOM CER is called for in this case.

These results, as before, show that the marginal RANDOM technique is more appropriate for estimating material cost CER's.

Comparisons Overall

We know that the cumulative data approach is not correct here. There is auto-correlation present in the data and this affects the estimated CER's greatly. These CER's were presented, then, for illustrative purposes and to point out some of the properties exhibited by the RANDOM CER's. This study does not recommend the cumulative cost approach for any airframe cost estimation.

In all cases the cumulative CER statistics seem superior to the respective marginal statistics. This is the result of the auto-correlation problem previously mentioned.

The cumulative RANDOM CER's show very small value for $\hat{\sigma}_\epsilon^2$ relative to $\hat{\sigma}_U^2$. This again can be a result of auto-correlated data. The

Table VIII
CER Coefficients and CER Statistics
Material Dollars

CER Type	Estimated Coefficients for the Explanatory Variables				Statistics			
	K	S	W	Q	R^2	$\hat{\sigma}_R^2$	$\hat{\sigma}_\epsilon^2$	$\hat{\sigma}_U^2$
MCR	3.7403 (1.24)	.7021 (1.13)	.6470 (2.39)	.7735 (28.53)	.998	-	.0109	.1320
MCT	4.0972 (2.95)	.1287 (.43)	1.0374 (7.24)	.7499 (10.31)	.918	.0969	-	-
MCL	7.0440 (2.00)	.3189 (.42)	.9506 (2.73)	-	.780	.1808	-	-
MMR	3.7288 (1.64)	.2505 (.51)	.9507 (4.26)	-.2204 (-6.81)	.995	-	.0324	.0694
MMT	3.6591 (3.06)	.1524 (.59)	1.0342 (8.37)	-.2331 (-5.03)	.853	.0713	-	-
MCUMT	3.8209 (1.55)	-.0110 (-.02)	1.0990 (4.49)	.8747 (2.57)	.929	.0830	-	-

marginal components are quite close to one another in magnitude in nearly all cases. The cumulative RANDOM CER components of variance indicate one other interesting fact, also. $\hat{\sigma}_{\epsilon}^2$ is small in comparison to $\hat{\sigma}_{U}^2$ due to auto-correlation, which in turn causes the value $\rho = \frac{\hat{\sigma}_{U}^2}{\hat{\sigma}_{U}^2 + \hat{\sigma}_{\epsilon}^2}$ to be large. Now, if ρ is very close to one, the large V matrix becomes very nearly singular. If V is singular, then the LAR or cumulative total cost technique is indicated. All this points out the inappropriateness of the cumulative data approach used by TT and indicates a different use of the data (marginal).

The marginal CER's, the only legitimate data approach used here, show very consistent quantities. The variance components associated with the RANDOM CER's tend to be much closer to one another in magnitude in every case except engineering hours. The nearly equal magnitudes for the component values in the other three categories does not favor either the LAR or TT technique. One also can see that in every category (excluding engineering again) the R^2 is better for the RANDOM CER as compared to both the marginal TT CER and the LAR CER. These two factors, then, indicate the marginal RANDOM technique is most appropriate.

One final comparison left to be considered in this chapter is between the cumulative total cost (CTC) approach and the LAR approach. The CTC approach was designed to eliminate the necessity for estimating observations as LAR did. As one can see, the results show that the cumulative total cost approach CER's are superior.

Prediction Analysis

This chapter will examine the predictive qualities of all the CER's developed. The cost to be predicted will be that of the F-14, lots one and two only. The comparison of predicted versus actual costs will be presented in Appendix D.

F-14 Explanatory Variables

The only variable that presents any problem here is the speed of the F-14. The speed is classified, but this paper can utilize any speed value published in any public record. The most recent published maximum speed for the F-14 was given as Mach 2.24. The method used to estimate this value in knots was by comparison to other similar aircraft with unclassified values given.

F-105 ---------- Mach 2.10 ----------1195 knots

F-4 ---------- Mach 2.27 ----------1220 knots

Interpolation between these two aircraft speed values yields an approximation or estimate of the maximum speed at best altitude for the F-14 of 1215 knots. Obviously, this estimate could cause prediction error but as was noted for the CER's presented, the speed explanatory variable was the least significant in all cases. This should minimize the effect of a poor speed estimate and result in acceptable predictions.

For each cost element there are three sets of predictions calculated. Set one includes the predictions for the first lot of F-14's using the RANDOM and Timson and Tihansky (TT) techniques for both the cumulative and marginal cost approaches. Set two includes the cumulative and

marginal predictions for the second lot of F-14's and for an additional comparison the RANDOM prediction will be in two steps: (1) the unadjusted prediction, and (2) the adjusted prediction (discussed later in this chapter). The third set of predictions will be the cumulative RANDOM and TT technique based on the prediction of cost at airframe quantity equal to 100. These values will be compared with the Large, et al. (LAR) based prediction. The reader is reminded that a LAR CER was not developed using marginal cost data. Table IX will summarize the explanatory values to be utilized for predictions.

Table IX
Explanatory Variable Values used for Prediction

Explanatory Variable	Value	
S	1215	
W	26500	
Q	12	Lot one cumulative
	38	Lot two cumulative
	5	Lot one marginal
	24*	Lot two marginal
	100	Large type predictions

*The value used for the material cost element predictions was 25 due to a significantly different learning curve slope estimation.

Adjusting Lot Two Predictions

One of the advantages to using the RANDOM technique is that previous information can be used to better predict the cost of a second or third, etc., lot of airframes. The adjustment factor is essentially some

function of the variance components multiplied by a function of the residual error values determined from earlier (in this case, lot cost) predictions and the observation of the actual values as they occur.

The development of the mathematical technique is a simple extension of the generalized least squares procedure as presented by Johnston (Ref 7:212-213). The model is

$$Y = XB + U \tag{34}$$

with the $E(U) = 0$ and $E(UU') = V$. The problem here is to predict a single value of the dependent variable y_o given the vector of explanatory variable values X_o. One can now write

$$y_o = X_o'B + u_o \tag{35}$$

where u_o is the true but unknown value of the disturbance. Again, as *before, assuming*

$$E(u_o) = 0 \tag{36}$$

$$E(u_o^2) = \sigma_o^2 = \sigma_R^2 \tag{37}$$

$$E(u_o\ U) = \begin{bmatrix} E(u_1\ u_o) \\ E(u_2\ u_o) \\ \vdots \\ E(u_n\ u_o) \end{bmatrix} = W \tag{38}$$

where W is the n x 1 vector of covariances of the prediction disturbance with the vector of sample disturbances.

Now Johnston defines a linear predictor as

$$\hat{p} = \hat{C}'\ Y \tag{39}$$

where $\hat{C}$ is a vector of n constants. If p is to be a best linear unbiased predictor, then one must choose $\hat{C}$ to minimize the predictor variance

$$\sigma_p^2 = E\left[(p - y_o)^2\right] \tag{40}$$

subject to $E(p-y_o) = 0$. With the use of Lagrange multipliers Johnston derives $\hat{C}$ as

$$\hat{C} = V^{*-1}\left[I - X(X'V^{*-1}X)^{-1}X'V^{*-1}\right]W + V^{*-1}X(X'V^{*-1}X)^{-1}X'V^{*-1}Y \tag{41}$$

which leads to

$$\hat{p} = \hat{C}'Y = X_o'\hat{B} + W'V^{*-1}E \tag{42}$$

where $E = Y - X'\hat{B}$ is the vector of GLS residuals and $X_o'\hat{B}$ is the unadjusted prediction.

In this paper the only prediction involving the above methods is the RANDOM technique prediction of lot two. In our case, for lot two,

$$V^* = \begin{bmatrix} & & 0 \\ & V & \vdots \\ & & 0 \\ 0\cdots 0 & & \hat{\sigma}_U^2 + \hat{\sigma}_\epsilon^2 \end{bmatrix} \tag{43}$$

and

$$W = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \hat{\sigma}_U^2 \end{bmatrix} \tag{44}$$

therefore,

$$\hat{P}_2 = \text{Adjusted Prediction Lot Two (RANDOM)}$$

$$= X_2' \hat{B} + \left(\frac{\hat{\sigma}_U^2}{\hat{\sigma}_U^2 + \hat{\sigma}_\epsilon^2} \right) e \qquad (45)$$

where e represents the actual cost of the first lot minus the predicted cost and $X_2' \hat{B}$ represents the unadjusted prediction for lot two. Of course, the adjusted prediction of lots three and on would require more complex matrix manipulation, but Eq (42) could be used with the proper construction of W, E, and V^{*-1}.

Prediction Intervals

Along with the actual point estimates, this chapter will also present prediction intervals. A user of a CER who wishes to predict a future observation usually wants or needs to make a statement regarding the confidence that can be placed in the prediction. There are two common measures of this type: confidence intervals and prediction intervals. Confidence intervals are estimated limits within which, with some specified probability, the mean of the distribution of all possible observations about the true regression line lies. Now, prediction intervals are the estimated limits within which, with some probability, the value of a single point estimate (future) lies. This study will use only the prediction interval confidence measure since there will be a comparison of the prediction capabilities of the three different types of CER's (Ref 13:2).

The actual equations used to obtain these intervals are developed in most statistical texts. Of course, the prediction interval formula for the RANDOM prediction is different from the other two prediction

formula and will be discussed later. The TT and LAR technique based CER's must use the following equation for the interval calculations for all predictions:

$$+/- \quad t_{(\alpha/2,\ n-k)}\ \hat{\sigma}_R \sqrt{1 + X_o'(X'X)^{-1}X_o} \qquad (46)$$

where $t_{(\alpha/2,\ n-k)}$ is the t statistic corresponding to the desired confidence level (in our case 90% confidence), $\hat{\sigma}_R$ is an estimate of the standard deviation of regression, X_o is the vector of explanatory variable values utilized to obtain the prediction, and X is the observation set of explanatory variable values used for regression. The above value is then added to and subtracted from the prediction to form a prediction interval.

The equation for calculating the interval for RANDOM lot one predictions is very similar to the previous one. The only change to Eq (46) is that the value V^{-1} is placed between the values, X'X:

$$+/- \quad t_{(\alpha/2,\ n-k)}\ \hat{\sigma}_R \sqrt{1 + X_o'(X'V^{-1}X)^{-1}X_o} \qquad (47)$$

where V is the matrix constructed of the variance components (see page 14).

Follow-on lot prediction intervals under the RANDOM technique are quite different. The prediction variance changes when more information becomes available. For example, in the case here, the prediction variance, σ_p^2, changes after the observance of the F-14 lot one cost. Essentially, the σ_p^2 must be re-estimated after each observation. The equations used to accomplish this re-estimation can be derived along the same lines as Johnston's development (Ref 7 :212-213). One must realize, though, that Johnston's equations for an updated prediction variance utilize an updated $\hat{B}$ vector. For example, in this study the

first lot results (cost observed) are added to the data set and then a new B vector could be estimated along with a new V^* matrix. Since this thesis did not update the $\hat{B}$ vector, the equations from Johnston must be modified.

Again, as in the adjustment of the F-14 lot two point prediction, W and V^* are defined as in Eqs (43) and (44). Now, using previous notation and further defining

y_1 = actual cost of lot 1

$\hat{p}_1$ = prediction of lot 1 cost

$$\rho = \frac{\hat{\sigma}_U^2}{\hat{\sigma}_U^2 + \hat{\sigma}_\epsilon^2}$$

X_1 = vector of explanatory variables associated with lot 1

the most obvious way to approach the development of the $\hat{\sigma}_{p_2}^2$ equation when one does not re-estimate the B vector is to begin with the equations resulting from Johnston's development concerning the adjustment of the lot two costs (Eq (45)). The equation is

$$\hat{p}_2 = X_2\hat{B} + \rho\, e \tag{48}$$

where

$$e = y_1 - X_1\hat{B} \tag{49}$$

and it was developed from Johnston's final result where given a vector $\hat{C}$ the best linear unbiased predictor is defined as

$$\begin{aligned}\hat{p} &= X_o\hat{B} + W'V^{*-1}E \\ &= \hat{C}Y\end{aligned} \tag{50}$$

Using Eqs (48) and (50) and the definition of $\hat{B}$ and e one can now obtain an equation for $\hat{C}$ that does not use an updated $\hat{B}$ vector for the $\sigma_{p_2}^2$

estimate associated with our lot two adjusted prediction:

$$\hat{B} = (X'V^{-1}X)^{-1} X'V^{-1}Y \tag{13}$$

e = error associated with the F-14 lot one prediction using the RANDOM CER

$$= y_1 - X_1 \hat{B} \tag{51}$$

Substituting into Eq (48)

$$\hat{p}_2 = (X_2 - \rho X_1) (X'V^{-1}X)^{-1} X'V^{-1}Y + \rho y_1 \tag{52}$$

Now, referring to Eqs (50) and (52), one can immediately determine a new $\hat{C}'$ vector for the case here with the original V and X matrices. It can be defined as the following row vector:

$$\hat{C}' = \left[(X_2 - \rho X_1)' (X'V^{-1}X)^{-1} X'V^{-1}Y \;\vdots\; \rho \right] \tag{53}$$

Remembering from Eq (40) the prediction variance is defined as

$$\sigma_p^2 = E\left[(p - y_o)^2 \right] \tag{40}$$

subject to $E(p-y_o) = 0$. Johnston minimizes σ_p^2 with the use of his $\hat{C}$ vector which uses the updated V^* and X matrices and for the F-14 lot two case would be

$$\hat{\sigma}_{p_2}^2 = \hat{C}'V^*\hat{C} + \hat{\sigma}_R^2 - 2\hat{C}'W \tag{54}$$

where $\hat{\sigma}_R^2$ is the estimated variance due to regression. Using the non-updated $\hat{C}$ vector developed here, one can now develop an equation for $\hat{\sigma}_{p_2}^2$ which accounts for the fact that in this thesis we did not re-estimate the coefficients, B, after observing the F-14 lot one acquisition cost.

$$\hat{\sigma}_{P_2}^2 = \hat{C}' V \hat{C} + \hat{\sigma}_R^2 - 2 \hat{C}' W$$

$$= \left[(X_2 - \rho X_1)(X' V^{-1} X)^{-1} X' \;\vdots\; \hat{\sigma}_U^2 \right] \hat{C}$$

$$+ \hat{\sigma}_U^2 + \hat{\sigma}_\epsilon^2 - 2\rho \hat{\sigma}_U^2$$

$$= (X_2 - \rho X_1)(X' V^{-1} X)^{-1} (X_2' - \rho X_1')$$

$$+ \hat{\sigma}_U^2 + \hat{\sigma}_\epsilon^2 - \rho \hat{\sigma}_U^2 \qquad (55)$$

Eq (55) is now in the proper form to be used in this thesis. The equation for the prediction variance for lot one, $\hat{\sigma}_{P_1}^2$, is Eq (55) with $\rho = 0$. Therefore, the lot two prediction variance will be significantly smaller than the variance associated with the F-14 lot one prediction. The interval should decrease dramatically. This, of course, gives the RANDOM technique a great advantage over the others when predicting follow-on lots of airframes.

All the intervals would be symmetric for the linear or exponential CER forms but in the case of the log-linear form used in this paper the interval is skewed so that the interval is larger on the high end than the low end. As was mentioned in an earlier chapter, this is more like the real world where there seems to be more cost overruns than underruns.

Prediction Presentation

Tables X, XI, XII, and XIII present the prediction results for engineering, tooling, labor and material cost elements, respectively. The costs presented are in thousands of hours/dollars and a percentage value is also given to facilitate interval width comparisons. This percent value is calculated by subtracting the lower or upper interval value from the point estimate (prediction) and then dividing this value by the point estimate. The value of this resulting calculation is then multiplied by 100 to obtain the percent interval.

Table X
Engineering Hour Predictions

CER Type/Lot #	Lower Limit	Percent	Prediction	Percent	Upper Limit
Cumulative Cost Data					
RANDOM Lot 1	6242	-30.3	8960	43.5	12860
TT Lot 1	6434	-29.7	9147	42.1	13002
RANDOM Lot 2 (Unadjusted)	8373	-29.8	11919	42.4	16968
RANDOM Lot 2 (Adjusted)	13526	-16.6	16210	19.8	19427
TT Lot 2	8856	-27.4	12104	37.7	16790
RANDOM 100 Airframes	10679	-29.5	15147	41.8	21483
TT 100 Airframes	11423	-26.4	15523	35.9	21095
LAR 100 Airframes	8373	-45.2	15280	82.5	27885
Marginal Cost Data					
RANDOM Lot 1	2708	-69.4	8837	226.3	28838
TT Lot 1	2779	-69.0	8978	223.0	29002
RANDOM Lot 2 (Unadjusted)	1377	-66.9	4160	202.1	12568
RANDOM Lot 2 (Adjusted)	1440	-66.4	4282	197.4	12737
TT Lot 2	1401	-66.5	4176	197.9	12442

Table XI
Tooling Hour Predictions

CER Type/Lot #	Lower Limit	Percent	Prediction	Percent	Upper Limit
Cumulative Cost Data					
RANDOM Lot 1	1973	-60.9	5050	155.9	12921
TT Lot 1	2556	-37.9	4117	61.1	6631
RANDOM Lot 2 (Unadjusted)	2853	-60.6	7248	154.0	18412
RANDOM Lot 2 (Adjusted)	6394	-22.0	8197	28.2	10509
TT Lot 2	4363	-35.2	6731	54.3	10383
RANDOM 100 Airframes	3879	-60.5	9816	153.1	24842
TT 100 Airframes	6711	-34.0	10169	51.5	15410
LAR 100 Airframes	6780	-39.5	11200	65.2	18501
Marginal Cost Data					
RANDOM Lot 1	1594	-73.5	6020	277.6	22731
TT Lot 1	1725	-64.4	4852	181.2	13645
RANDOM Lot 2 (Unadjusted)	886	-72.2	3192	260.3	11502
RANDOM Lot 2 (Adjusted)	1152	-63.0	3109	169.8	8388
TT Lot 2	1085	-61.8	2838	161.6	7425

Table XII
Labor Hour Predictions

CER Type/Lot #	Lower Limit	Percent	Prediction	Percent	Upper Limit
		Cumulative Cost Data			
RANDOM Lot 1	3928	-44.0	7013	78.5	12521
TT Lot 1	4077	-39.2	6705	64.4	11026
RANDOM Lot 2 (Unadjusted)	7979	-43.7	14173	77.6	25176
RANDOM Lot 2 (Adjusted)	9338	-15.3	11030	18.1	13028
TT Lot 2	8875	-36.4	13951	57.2	21930
RANDOM 100 Airframes	14439	-43.6	25583	77.2	45325
TT 100 Airframes	16725	-35.2	25806	54.3	39819
LAR 100 Airframes	12537	-48.9	24543	95.8	48048
		Marginal Cost Data			
RANDOM Lot 1	4143	-44.0	7396	78.5	13203
TT Lot 1	4258	-39.6	7048	65.5	11666
RANDOM Lot 2 (Unadjusted)	4957	-42.5	8614	73.8	14967
RANDOM Lot 2 (Adjusted)	4693	-36.5	7396	57.6	11654
TT Lot 2	5273	-37.1	8389	59.1	13346

Table XIII
Material Dollar Predictions

CER Type/Lot #	Lower Limit	Percent	Prediction	Percent	Upper Limit
		Cumulative Cost Data			
RANDOM Lot 1	14935	-51.3	30674	105.4	63001
TT Lot 1	19972	-46.8	37526	87.9	70511
RANDOM Lot 2 (Unadjusted)	36740	-50.9	74818	103.6	152361
RANDOM Lot 2 (Adjusted)	125174	-22.0	160531	28.2	205878
TT Lot 2	50198	-43.6	89076	77.4	158065
RANDOM 100 Airframes	77982	-50.7	158145	102.8	320717
TT 100 Airframes	106185	-42.3	184034	73.3	318959
LAR 100 Airframes	69801	-60.5	176830	153.3	447968
		Marginal Cost Data			
RANDOM Lot 1	18204	-45.3	33294	82.9	60893
TT Lot 1	20943	-41.0	35482	69.4	60117
RANDOM Lot 2 (Unadjusted)	28219	-44.2	50597	79.3	90723
RANDOM Lot 2 (Adjusted)	56668	-32.6	84059	48.3	124690
TT Lot 2	32592	-38.3	52828	62.1	85628

The marginal results presented are the CER predictions multiplied by the appropriate number of airframes included in the lot. Therefore, even though marginal refers to average cost per airframe per lot, the tables will show a total lot cost for better comparison qualities.

A 90 percent prediction interval was calculated for the predictions utilizing the Student-t value of 1.699 for the non-normalized airframe quantity CER's (degrees of freedom = 29), and 1.943 for the LAR CER's (degrees of freedom = 6). Cumulative total CER predictions and prediction intervals were not calculated here because of a lack of complete F-14 cost data.

The average percent intervals under the marginal, cumulative and LAR cost approaches over all cost categories are listed in Table XIV. The intervals are large in all cases. The TT prediction intervals are the smallest when predicting lot one costs and the cost of 100 airframes. But, the RANDOM type prediction interval decreases greatly when predicting the lot two cost. Of course, this occurs because the RANDOM interval equations take into account the fact that the lot one cost is known. This decreased interval is comparable or better than the TT interval for lot two in both the marginal and cumulative cost approaches.

One may argue here that the TT intervals could have been reduced by re-estimating the B coefficients after the lot one cost was known. But, remember that this was not done in the RANDOM case as explained earlier. There is no reason to believe that had the coefficients been re-estimated under both techniques, the comparison would not have shown the same comparison results. This is, of course, because the RANDOM technique assumes there are two components of variance and when the lot one cost

Table XIV
Average Percent Intervals

CER Type/Lot #	Lower Limit	Upper Limit
	Cumulative Cost Data	
RANDOM Lot 1	-46.6	95.8
TT Lot 1	-38.4	63.9
RANDOM Lot 2 (Adjusted)	-19.0	23.6
TT Lot 2	-35.7	56.7
	Quantity = 100	
RANDOM	-46.1	93.7
TT	-34.5	53.8
LAR	-48.5	99.2
	Marginal Cost Data	
RANDOM Lot 1	-58.1	166.3
TT Lot 1	-53.5	134.8
RANDOM Lot 2 (Adjusted)	-49.6	118.3
TT Lot 2	-50.9	120.2

is observed one is actually assuming that σ_U^2 is being observed thereby reducing the total variance associated with the lot two prediction. This is what the RANDOM technique with the adjustments made for the lot two actually does and the results here show a better average 90% prediction interval for lot two especially under the cumulative cost approach. One can see that there is no decisively superior technique in the marginal lot two intervals presented. This may be a consequence of the possible auto-correlation problem discussed earlier in the cumulative CER's. The marginal data approach eliminates this problem if it is present and may be showing more realistic CER results. This also accounts for the fact that in all cases here except the material cost category the cumulative intervals are narrower than the marginal intervals.

These results are consistent with the hypotheses presented in Chapter I. That is, the TT technique underestimates the variance for predicting the cost of a new type airframe (the TT lot one intervals are smaller than RANDOM's) and overestimates the variance for predicting the cost of a follow-on lot of airframes (the TT lot two intervals are larger than RANDOM's). Also, the LAR intervals are the largest of the three Q = 100 airframe predictions as was hypothesized in Chapter I.

Before the reader draws any conclusions on the basis of the previous chapter and this one, he is advised to refer to Appendix D where the point predictions themselves are compared to actual cost (this is privileged information).

The reader is again reminded that the cumulative results are suspect due to the auto-correlation that is very likely present.

VI. Summary and Conclusions

The RANDOM modeling technique is adaptable to airframe cost estimation with the use of the unbiased, maximum likelihood variance component estimators put forth by Searle and developed in Appendix A.

The RANDOM CER coefficients along with those of the redeveloped Large, et al. (LAR) and Timson and Tihansky (TT) type CER's (all presented in Chapter IV) are comparable to the coefficients estimated in past studies using the same explanatory variables.

The cumulative data approach looks much better than the marginal approach in almost all comparisons made in Chapters IV and V and Appendix D. But, this approach is inappropriate. The cumulative approach has the inherent problem of auto-correlated data. This factor is known to inflate R^2 values and affects the estimated variance component values for the cumulative RANDOM CER's, i.e., the component associated with different airframes is much larger than the other. These problems and the correlation assumed to be present in lot data of like airframes all point to the LAR type CER as being the most applicable when compared to the TT and RANDOM techniques under the cumulative data approach.

However, the LAR type CER's are not appropriate either as one notices the comparison of them with the cumulative total cost CER's. This other technique is very much like LAR's except that the total quantity ever produced of a particular airframe type is included as a variable to predict total cumulative cost. This indicates that even though the LAR technique eliminates both lot correlation and auto-correlation problems, there still exists the "observation" estimation

factor. This is why the cumulative total CER R^2's are superior to the LAR R^2's in all cases. The addition of the airframe quantity variable eliminated the need for learning curve estimation and resulted in a better CER.

Thus, if the cumulative data approach must be utilized, the most appropriate technique is the cumulative total cost. However, as discussed in Chapters III and IV, the LAR and cumulative total type techniques are throwing valuable information away that is contained in the separate lot data. This, of course, brings us to the marginal type CER's. The RANDOM estimated variance components presented in Chapter IV do not point to either the TT or LAR technique as being more appropriate. A more general technique like RANDOM is indicated. The R^2 values add more weight to the above statement; the RANDOM R^2's are best in all but the engineering marginal type CER's.

The *F*-14 prediction results presented in Chapter V and Appendix D lead to similar findings from the comparison of the marginal and cumulative CER's. That is, the cumulative prediction intervals look much tighter than the marginal ones when in fact they are incorrect. Also, the cumulative point predictions look better than the marginal predictions. But, again, auto-correlation is the problem here. It causes the cumulative prediction intervals to look smaller than they really are.

The point prediction comparisons seem to favor the cumulative approach also. This is especially true of the lot two predictions. However, one must remember that the cumulative lot two predictions must be adjusted for proper comparisons with the same marginal predictions. The adjustment does enlarge the cumulative lot two prediction error in most cases.

These prediction results do indicate that our hypotheses are generally correct. The marginal lot one intervals for the RANDOM predictions are larger than the respective TT intervals. We did hypothesize that the TT technique underestimates the prediction variance for a new type airframe. However, the hypothesis that the TT technique overestimates the prediction variance of follow-on lots is not very evident in the marginal lot two interval results. The RANDOM and TT intervals are very nearly equal. Also, the LAR prediction intervals are larger than necessary.

The point prediction results are not explained by our hypotheses. The RANDOM predictions should be superior especially for lot two costs. This did not occur in this study. This could be the result of an inherent problem in the RANDOM lot two adjustment process. That is, when the RANDOM prediction errors are of opposite signs for lots one and two (unadjusted), the adjustment process actually increases the lot two prediction error. This did occur in two of the four cost categories in this study. This caused the RANDOM prediction to appear inferior. But, this particular problem should disappear when predicting lots three, four, etc. The RANDOM predictions and associated intervals should become much better than the same TT type results as the number of observed lots increases. One must also realize that this is a very small test of the predictive capabilities. The F-14 data may be inaccurate or atypical. A great deal more analyses in this area must be accomplished.

This study laid the foundation for a very promising "new" technique for estimating airframe acquisition cost models. This technique is much more general than present methods and this characteristic can be utilized to determine which of the two most used airframe cost estimating

techniques, Large's or Timson and Tihansky's, is more appropriate. If the LAR or TT techniques can not be justified by an analysis of the RANDOM variance components estimated, then the RANDOM technique itself may be appropriate. Unfortunately, the prediction results were inconsistent as compared to the CER statistics. This is not surprising when one uses only one airframe type and two lots as a test case.

Even with some contradictory results the RANDOM technique is very promising. This is especially evident in the CER comparisons contained in Chapters IV and V where the marginal RANDOM CER consistently proved to be the most appropriate. There are obvious shortcomings in the TT and LAR techniques. The RANDOM marginal technique is able to account for these shortcomings and therefore is an important additional tool to be used in airframe cost estimation. One could not conclude from this small study that the RANDOM technique is the most appropriate technique for all airframe cost estimation. But, it does indicate that the LAR and TT techniques are definitely not appropriate in all cases. This new technique deserves the interest of DoD.

New Research

What new directions of research are indicated as a result of this study? The most obvious "next step" would be to conduct a very similar study using all the airframe cost data available and then determine the best explanatory variables to use for each CER estimated with some sort of step-wise procedure. The same format could be used as was in this thesis, except this new study should do two things differently. First, the coefficients should be re-estimated after observing a previous

lot cost. This should result in better prediction capability analysis for all techniques utilized. Second, the engineering cost element data problem should be investigated. There may be a way to account for the heavy non-recurring cost loading in the first lots. The marginal RANDOM technique proved to be the most appropriate in many comparisons presented in this thesis, but this new study must be accomplished to validate these findings.

As a separate thesis effort or part of the above mentioned study, the significance level of the estimated variance components should be addressed. The distributional property of the components has to be discovered and if that is done, then one could test the hypotheses that either component is zero or not. If $\hat{\sigma}_U^2 = 0$, then use the TT approach; if $\hat{\sigma}_\epsilon^2 = 0$, use the LAR approach; and if neither is equal to zero, then the RANDOM technique may be the method prescribed to estimate airframe costs.

Bibliography

Bibliography

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

2. Cunningham, E. P. and C. R. Henderson. "An Iterative Procedure for Estimating Fixed Effects and Variance Components in Mixed Model Situations." Biometrics, 24:13-25 (1968).

3. Eisenhart, Churchill, ed. Techniques of Statistical Analysis. New York: McGraw-Hill Book Co., 1947.

4. Handel, Vernon. 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.

5. Henderson, C. R. "Estimation of Variance and Covariance Components." Biometrics, 9:229-252 (1953).

6. Hogben, D., et al. OMNITAB II User's Reference Manual. Washington, D. C.: National Bureau of Standards, 1971.

7. Johnston, J. Econometric Methods. New York: McGraw-Hill Book Company, Inc., 1972.

8. Large, Joseph P., et al. Parametric Equations for Estimating Aircraft Airframe Costs. RAND Report R-1693-PA & E. Santa Monica, California: The RAND Corporation, May 1975.

9. Levenson, G. S. and S. M. Barro. Cost-Estimating Relationships for Aircraft Airframes. RAND Report RM-4845-PR. Santa Monica, California: The RAND Corporation, February 1966.

10. Levenson, G. E., et al. Cost Estimating Relationships for Aircraft Airframes. RAND Report R-761-PR. Santa Monica, California: The RAND Corporation, December 1971.

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

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

13. Timson, F. S. and D. F. Tihansky. Confidence in Estimating Airframe Costs: Uncertainty Assessments in Aggregate Predictions. RAND Report R-903-PR. Santa Monica, California: The RAND Corporation, October 1972.

14. Theil, Henri. Principles of Econometrics. New York: John Wiley & Sons, Inc., 1971.

15. Thompson, R. "Iterative Estimation of Variance Components for Non-Orthogonal Data." Biometrics, 25:767-773 (1969).

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

Data Source

Data Source

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

Supplementary Bibliography

Supplementary Bibliography

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

2. Dunne, William 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.

3. Mendenhal, William. The Design and Analysis of Experiments. Belmont, California: Wadsworth Publishing Company, Inc., 1968.

4. Neter, J. and W. Wasserman. Applied Linear Statistical Models. Homewood, Illinois: Richard D. Irwin, Inc., 1974.

5. Noah, J. W., et al. Estimating Aircraft Acquisition Costs by Parametric Methods. J. Watson Noah Associates Report FR-103-USN (Abridged), September 1973.

6. O'Brien, Patrick W. Joint Generalized Least Squares Applied to Cost Estimation for Fighter Aircraft. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, December 1974.

7. Scheffe', Henry. The Analysis of Variance. New York: John Wiley & Sons, Inc., 1959.

8. Searle, S. R. "Another Look at Henderson's Methods of Estimating Variance Components." Biometrics, 24:749-800 (1968).

9. Zambreka, P. Frontiers in Econometrics. New York: Academic Press, 1974.

Appendix A

The Development of the Variance Component Estimation Procedure

Appendix A

The Development of the Variance Component Estimation Procedure

This section will first discuss the general application of mixed models and the variance components associated with them. Then a detailed development of the variance estimation process utilized in this study will be presented.

Mixed Fixed Effects/Random Effects Model

Eisenhart introduced the term "mixed model" to describe the models used for experiments where some of the effects, such as animal effects, can be random effects while other effects, such as different treatments, are regarded as fixed (Ref 3:1-21). The method of estimation of the variance components is fully understood for various mixed data arrangements when there is a high degree of symmetry. However, for a general non-orthogonal design there are difficulties and no simple, known, method is optimal under all conditions (Ref 15:767).

The lack of balance in a data set is a problem encountered often in the real world. This lack of balance causes the data to be non-orthogonal. In this study the data is in fact unbalanced, i.e., there are different numbers of lots associated with each type airframe. This non-orthogonality of the lot data necessitates a very specialized method to estimate the variance components.

Cunningham and Henderson proposed a general method for estimating the variance components for an unbalanced situation with iterative

calculations (Ref 2 :13-25). But Thompson discovered an algebraic oversight on their part which caused the iterative process to converge to unreasonable estimates (Ref 15:767). Thompson corrected this error and Searle further refined the technique to fit the situation when one has a mixed model having one random factor (Ref 12:465-470).

The above procedure given by Searle is tailor made for the RANDOM technique described in this study. Of course, the one random factor here is the type airframe, while the fixed portion of the model is associated with the aircraft characteristics.

Development of the Estimation Procedure

This section will parallel the development of the estimation technique as presented by Searle (Ref 12:465-470).

To begin with, the model must be defined as

$$Y = XB + ZU + \in \tag{56}$$

with the rank of X equal to r(r(X)= r), with B representing $g \geq r$ fixed effects and U, in representing the random effects, contains t effects for one random factor (airframe), having a variance equal to σ_U^2. Z then has full column rank, t, with the columns summing to one. Therefore,

$$r(X\ Z) = r(X) + t - 1 = r + t - 1 \tag{57}$$

where the space between X and Z indicates a partitioned matrix. Also, as one can see, the matrix Z'Z is diagonal and non-singular.

Since the model is mixed, the normal estimation procedure here would be to use the fitting constants method (Henderson's Method 3) (Ref 5:226-252), where the sum of squares of error (SSE) and the sum of squares due to regression, R(_), are defined as follows.

$$SSE = Y'Y - R(B,U) \tag{58}$$

$$R(U|B) = R(B,U) - R(B) \tag{59}$$

with

$$E(SSE) = [N - (r + t - 1)]\,\sigma_\epsilon^2 \tag{60}$$

Additionally, from the fitting constants method

$$E[R(U|B)] = \sigma_U^2 \operatorname{tr}[Z'Z - Z'X(X'X)^- X'Z] + \sigma_\epsilon^2 [r(X) + t - 1 - r(X)] \tag{61}$$

Thus, point estimates of the variance components are

$$\hat{\sigma}_\epsilon^2 = \frac{Y'Y - R(B,U)}{N - r(X) - t + 1} \tag{62}$$

and

$$\hat{\sigma}_U^2 = \frac{R(U|B) - \hat{\sigma}_\epsilon^2 (t - 1)}{\operatorname{tr}[Z'Z - Z'X(X'X)^- X'Z]} \tag{63}$$

Now, a computational difficulty exists in the preceding formulation. That is

$$R(B,U) = Y'[X\ Z] \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix}^- \begin{bmatrix} X' \\ Z' \end{bmatrix} Y \tag{64}$$

This calculation can be very involved due to the large size of Z. The "absorption process" described by Searle permits easier calculation as follows (Ref 12:266-269): With

$$R(U) = Y'Z(Z'Z)^{-1} Z'Y \tag{65}$$

one finds that

$$R(B|U) = R(B,U) - R(U) \tag{66}$$

simplifies with the substitution from Eqs (64) and (65), to

$$R(B|U) = B^{o}X' \left[I - Z(Z'Z)^{-1} Z'\right] Y \tag{67}$$

where

$$B^{o} = Q^{-} X' \left[I - Z (Z'Z)^{-1} Z'\right] Y \tag{68}$$

and

$$Q = X'X - X'Z (Z'Z)^{-1} Z'X \tag{69}$$

From the above simplification one can now use Eqs (65) and (67) to calculate

$$R(B,U) = R(B|U) + R(U) \tag{70}$$

$$R(U|B) = R(B|U) + R(U) - R(B) \tag{71}$$

where

$$R(B) = Y'X(X'X)^{-} X'Y \tag{72}$$

The crucial result derived from the above equations is that B^{o} of Eq (68) is a solution to

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix} \begin{bmatrix} B^{o} \\ U^{o} \end{bmatrix} = \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix} \tag{73}$$

which actually are the ordinary least squares equations for B^{o} and U^{o} assuming that U is a vector of fixed rather than random effects.

For one to develop the comparable equations for obtaining the maximum likelihood solutions for the fixed effects requires more detail than is warranted here. Searle develops the theory and the results

are that (Ref 12:461-462)

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda I \end{bmatrix} \begin{bmatrix} B^* \\ U^* \end{bmatrix} = \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix} \tag{74}$$

where

$$\lambda = \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_U^2} \tag{75}$$

This is where Thompson picked up an error in the presentation by Cunningham and Henderson. They falsely deduced that since Eq (73) is the same as Eq (74) except for the quantity, $(Z'Z + \lambda I)$, replacing $Z'Z$, one could make this replacement throughout the entire variance component estimation process described by Eqs (63) through (72). The result is an iterative procedure based on the maximum likelihood equations implicit in Eq (74). Therefore, the variance component estimating equations become

$$\sigma_\epsilon^{*2} = \frac{Y'Y - R^*(B,U)}{N - r(X) - t + 1} \tag{76}$$

and

$$\sigma_U^{*2} = \frac{R^*(U|B) - \sigma_\epsilon^{*2}(t-1)}{tr\left[Z'Z + \lambda I - Z'X(X'X)^- X'Z\right]} \tag{77}$$

The comparable $R^*(-)$terms are

$$R^*(B,U) = R^*(B|U) + R^*(U) \tag{78}$$

$$R^*(U) = Y'Z\, P^{-1}\, Z'Y \tag{79}$$

$$R^*(B) = R(B) = Y'X(X'X)^- X'Y \tag{80}$$

and

$$R^*(U|B) = R^*(B|U) + R^*(U) - R^*(B) \tag{81}$$

where

$$P = Z'Z + \lambda I \tag{82}$$

Thompson, in his article referenced earlier, pointed out that the error here is the false assumption that the expected values of SSE* and $R^*(U|B)$ are the same as those of SSE and $R(U|B)$ shown in Eqs (60) and (61). This is not so and consequently the estimators given by Eqs (76) and (77) are not unbiased.

Now to derive maximum likelihood unbiased estimates of the components Thompson modified the equations by noticing first that from Eq (82)

$$\begin{aligned} P^{-1}Z'(ZZ'\,\sigma_U^2 + \sigma_\epsilon^2 I) &= P^{-1}(Z'Z\,\sigma_U^2 + \sigma_\epsilon^2 I)\,Z' \\ &= P^{-1}(Z'Z + \lambda I)\,\sigma_U^2\,Z' \\ &= P^{-1}\,P\,Z'\,\sigma_U^2 \\ &= Z'\,\sigma_U^2 \end{aligned} \tag{83}$$

second, from the model (Eq (56)),

$$E(YY') = XBB'X' + ZZ'\,\sigma_U^2 + \sigma_\epsilon^2 I \tag{84}$$

and finally, using the properties of the trace operator, the expected value of $R^*(U)$ is

$$E\left[R^*(U)\right] = \mathrm{tr}\left[ZP^{-1}Z'XBB'X' + ZZ'\,\sigma_U^2\right] \tag{85}$$

Next, defining the variable T

$$T = I - ZP^{-1}Z' \tag{86}$$

Eq (83) results in

$$T(ZZ' \sigma_U^2 + \sigma_\epsilon^2 I) = \sigma_\epsilon^2 I \tag{87}$$

so that now

$$E\left[R^*(B|U)\right] = tr\left[TXBB'X' + TX(X'TX)^{-}X' \sigma_\epsilon^2\right] \tag{88}$$

and

$$E\left[R^*(B)\right] = tr\left[XBB'X' + X(X'X)^{-}X'(ZZ' \sigma_U^2 + \sigma_\epsilon^2 I)\right] \tag{89}$$

Now, from Eq (81) and using Eqs (85), (88), and (89)

$$\begin{aligned} E\left[R^*(U|B)\right] &= \sigma_U^2 \, tr\left[Z'Z - Z'X(X'X)^{-}X'Z\right] \\ &\quad + \sigma_\epsilon^2 \, tr\left[X'TX(X'TX)^{-} - X'X(X'X)^{-}\right] \end{aligned} \tag{90}$$

But, the rank of XTX' equals the rank of X and

$$tr\left[X'X(X'X)^{-}\right] = r(X) \tag{91}$$

making the last term of Eq (90) equal to zero so that

$$E\left[R^*(U|B)\right] = \sigma_U^2 \, tr\left[Z'Z - Z'X(X'X)^{-}X'Z\right] \tag{92}$$

Also, from the above equations one can derive

$$\begin{aligned} &E\left[Y'Y - R^*(B,U)\right] \\ &\quad = E\left[Y'Y - R^*(U) - R^*(B|U)\right] \\ &\quad = \left[N - r(X)\right] \sigma_\epsilon^2 \end{aligned} \tag{93}$$

where N is the total number of observations.

This all results, then, in an iterative procedure utilizing $\lambda = \hat{\sigma}_\epsilon^2/\hat{\sigma}_U^2$ to obtain the unbiased, maximum likelihood estimators of σ_ϵ^2 and σ_U^2 using the equation

$$\hat{\sigma}_\epsilon^2 = \frac{Y'Y - [R^*(U) + R^*(B|U)]}{N - r(X)}$$

$$= \frac{[Y'Y - R^*(B,U)]}{N - r} \qquad (94)$$

and

$$\hat{\sigma}_U^2 = \frac{R^*(U) + R^*(B|U) - R^*(B)}{tr\,[Z'Z - Z'X(X'X)^-X'Z]}$$

$$= \frac{R^*(U|B)}{tr\,[Z'Z - Z'X(X'X)^-X'Z]} \qquad (95)$$

The variance component estimation process, therefore, is accomplished by taking an initial value of λ, calculating P and T, which leads to the calculation of the necessary terms $R^*(U)$, $R^*(B|U)$, etc., and finally the variance components themselves. Then λ is again calculated with the new variance estimators and the procedure is repeated until convergence.

The Equations Summarized

This study utilized the equations developed in the previous section in the following manner. First of all an initial estimate of the variance components was obtained by using Henderson's Method 3. The equations programmed using the OMNITAB language and the symbols defined earlier were

$$T_o = Y'Y \qquad (96)$$

$$M = X'[I - Z(Z'Z)^{-1} Z'] \tag{97}$$

$$Q = MX \tag{98}$$

$$B^{o} = Q^{-}MY \tag{99}$$

$$R(B|U) = B^{o\prime}MY \tag{100}$$

$$R(U) = Y'Z(Z'Z)^{-1} Z'Y \tag{101}$$

$$R(B) = Y'X(X'X)^{-} X'Y \tag{102}$$

$$R(B,U) = R(B|U) + R(U) \tag{103}$$

$$R(U|B) = R(B,U) - R(B) \tag{104}$$

and

$$c = tr\left[Z'Z - Z'X(X'X)^{-} X'Z\right] \tag{105}$$

Now with the above values calculated

$$\hat{\sigma}_{\epsilon}^{2} = \frac{\left[T_{o} - R(B,U)\right]}{N - r - t + 1} \tag{106}$$

and

$$\hat{\sigma}_{U}^{2} = \frac{\left[R(U|B) - (t-1)\,\hat{\sigma}_{\epsilon}^{2}\right]}{c} \tag{107}$$

The above first estimates were used to calculate the initial value, λ, for the first iteration of the process summarized as follows:

$$\lambda = \frac{\hat{\sigma}_{\epsilon}^{2}}{\hat{\sigma}_{U}^{2}} \tag{108}$$

$$P = Z'Z + \lambda I \tag{109}$$

$$T = I - ZP^{-1}Z' \tag{110}$$

$$R^{*}(B|U) = Y'TX(X'TX)^{-}X'TY \tag{111}$$

$$R^{*}(U) = Y'ZP^{-1}Z'Y \tag{112}$$

$$R^*(B) = R(B) = Y'X(X'X)^{-}X'Y \tag{113}$$

$$R^*(B,U) = R^*(B|U) + R^*(U) \tag{114}$$

and

$$R^*(U|B) = R^*(B,U) - R^*(B) \tag{115}$$

The estimates then are

$$\hat{\sigma}_{\epsilon}^2 = \frac{[T_o - R^*(B,U)]}{N - r} \tag{116}$$

and

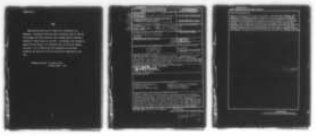
$$\hat{\sigma}_{U}^2 = \frac{R^*(U|B)}{c} \tag{117}$$

With the new estimates of σ_{ϵ}^2 and σ_{U}^2 a new value for λ is calculated and Eqs (109) through (117) are reaccomplished and so on until convergence.

AD-A032 627 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO SCH--ETC F/G 5/1
AIRCRAFT AIRFRAME COST ESTIMATION UTILIZING A COMPONENTS OF VAR--ETC(U)
OCT 76 R C MARCOTTE

UNCLASSIFIED GOR/SM/76D-10 NL

2 OF 2
AD
A032627



END

DATE
FILMED
1-77

Vita

Ronald Marcotte was born on 8 March 1947 in Manchester, New Hampshire. He graduated from high school in Sanford, Maine in 1965 and then attended the United States Air Force Academy where he received a Bachelor of Science degree in June 1969. He completed pilot training in August 1970 and became a T-37 Instructor Pilot at Moody AFB, Georgia. He served a tour in Thailand and the Phillipines as an aircraft controller and entered the Air Force Institute of Technology in June 1975.

Permanent address: 29 Farview Drive
Sanford, Maine 04073

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER GOR/SM/76D-10	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AIRCRAFT AIRFRAME COST ESTIMATION UTILIZING A COMPONENTS OF VARIANCE MODEL.		5. TYPE OF REPORT & PERIOD COVERED M. S. Thesis 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Ronald C. Marcotte Captain, USAF		8. CONTRACT OR GRANT NUMBER(s) Master's thesis,
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 October 1976 13. NUMBER OF PAGES 116
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) 12 98p.		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited. Appendixes B, C, and D are limited to U. S. Government agencies only; proprietary information; October 1976. Other requests for Appendixes (B, C, and D) of this document must be referred to Dean of Engineering, Air Force Institute of Technology (AFIT/EN), Wright-Patterson Air Force Base, Ohio 45433.

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

Approved for public release; IAW 190-17

JERRAL F. GUESS, Captain, USAF
Director of Information

18. SUPPLEMENTARY NOTES

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

Parametric Cost Estimation
Aircraft Airframe
Components of Variance Models
Parametric Airframe Cost Estimation
Airframe Cost Estimation
Aircraft Cost Estimation

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

Previous studies into airframe acquisition cost estimation do not explicity recognize the existence of correlation in the historical data. If one believes this data problem exists, then it is possible to develop a components of variance model that takes the problem into account. It is a more general model that recognizes two sources of error: (1) error due to different types of airframes and (2) overall or ordinary regression error. The variance of these two errors can be estimated and then can be utilized along with the technique of generalized least squares to obtain a cost estimating relationship which

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

012225
AB

Block 20 (continuation)

explicitly accounts for the data correlation. This modeling technique, when compared to techniques presently in service, shows that present estimating relationships underestimate the variance of the cost prediction of a new type airframe and overestimate the variance of the cost prediction of a follow-on airframe. Also, those existing techniques which implicitly recognize data correlation do not make use of all the data information available and therefore produce estimates with poor confidence/prediction intervals. The modeling technique developed here is an improvement over the present technique utilized and advances the state of the art of parametric airframe cost estimation greatly.