FITTING THE EQUATION $Z = a + bX + cY$ WHEN ALL VARIABLES ARE SUBJECT TO ERROR

A Thesis

Presented to

the Faculty of the Department of Industrial Engineering

The University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Industrial Engineering

by

Chung- Yung Chang

May 1973

ACKNOWLEDGEMENTS

The author wishes to express his deepest gratitude to Dr. B. T. Rhodes for his valuable suggestions and encouraging guidance towards the completion of this thesis. He is also indebted to Dr. C. E. Donaghey for his advice in the FORTRAN IV programs.

FITTING THE EQUATION $Z = a + bX + cY$ WHEN ALL VARIABLES ARE SUBJECT TO ERROR

An Abstract of a Thesis
Presented to
the Faculty of the Department of Industrial Engineering
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Industrial Engineering

by
Chung-Yung Chang
May 1973

ABSTRACT

In some cases of linear regression analysis, there exists a functional relationship $Y = f(X_1 , \ldots , X_n)$ among variables, but all variables cannot be observed because of measurement error. Instead of observing $Y, X_1, \ldots , X_n$, we obtain $y, x_1, \ldots , x_n$, where $y = Y + e_y$, $x_1 = X_1 + e_{x1}, \ldots , x_n = X_n + e_{xn}$. $e_y, e_{x1}, \ldots , e_{xn}$ are random errors with means 0 and variances $\sigma_y^2 , \sigma_{x1}^2 , \ldots , \sigma_{xn}^2$. In this case where all variables are subject to error, the only equation studied in the literature contains only one dependent variable and one independent variable; $Y = a + bX$. In 1941, A. Wald gave a simple method which could estimate the parameters of the equation $Y = a + bX$ by dividing the data into two groups for estimating the slope and fitting a line paralleling to this slope through the mean point $(\bar{x} , \bar{y})$ of all observations. In 1949, M. S. Bartlett found another simple method by dividing the data into three groups and using the two extreme groups for fitting the slope. In this thesis the techniques of both these simple methods will be extended to a multiple linear regression model. The linear model studied here is

$$Z = a + bX + cY$$

The purpose of this thesis is to make consistent point estimates of the parameters of the above equation by extending Wald's method and Bartlett's method. Proofs of consistency are given.

A simulation study was used to compare these two methods to the least squares method. The results indicate that all three methods give similar results. In addition it appears that all three methods are biased and that the bias is a function of the variances of the random errors.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

In today's science, it is of interest to describe and predict events in the world in which we live. One way to accomplish this goal is by examining the effect that some variable quantities, or events, exert on others and finding an equation that relates these quantities or events in the real world. In many areas of scientific endeavor, the relationship of variables may possibly be expressed as functional equation $Y = f(X_1, \ldots ,X_n)$.

When the function f or an approximation to f is obtained, it can be used to predict Y. Regression analysis is a procedure to estimate the parameters of an unknown equation from data. Usually in the linear regression analysis one assumes that the independent variables of the equation $Y = f(X_1, \ldots , X_n)$ can be measured or observed, but the dependent variable cannot be measured exactly. Instead of observing Y we actually observe y where $y = Y + e_y$. Generally e_y is random error (or normal random error) with zero mean and σ_y^2 variance. In this case the least squares method or the maximum likelihood method can be used to estimate the parameters.

In some cases even though there exists a functional relationship $Y = f(X_1, \ldots X_n)$ among variables, all variables cannot be observed because of error of measurement. Instead of observing $Y, X_1, \ldots X_n$, we obtain $y, x_1, \ldots x_n$, where $y = Y + e_y$, $x_1 = X_1 + e_{x1}$,

... , $x_n = X + e_{xn}$. The set $\{e_y, e_{x1}, \dots , e_{xn}\}$ are random errors with zero means and $\sigma_y^2, \sigma_{x1}^2, \dots , \sigma_{xn}^2$ variances. This case violates the assumption generally used in linear regression analysis, therefore presenting complexities that cannot be handled by the least squares method or the maximum likelihood method without additional assumptions.

In the case where all variables are subject to error, the only equation studied so far contains only one dependent variable and one independent variable; $Y = a + bX$ where Y and X will be unobservable quantities. Actually y and x are obtained where $y = Y + e_y$ and $x = X + e_x$, and e_y and e_x are random errors with zero means and σ_y^2 and σ_x^2 variances. Thus the only model studied in the literature is

$$y - e_y = a + b(x - e_x) \qquad (1\text{-}1)$$

In the case where e_y and e_x are normal and the ratio $\lambda = \sigma_x^2/\sigma_y^2$ is known, the maximum likelihood method for estimating the parameters a, b, σ_x^2 , and σ_y^2 in equation (1-1) is given in Graybill [1]. If the ratio is unknown, there are two simple methods which can handle this problem. In 1941, A. Wald [2] gave a simple method which could estimate the parameters of the equation $Y = a + bX$ by dividing the data into two groups to estimate the slope and fitting a line paralleling to this slope through the mean point $(\bar{x} , \bar{y})$ of all observations. In 1949, M. S. Bartlett [3] found a more efficient way by dividing the data into three groups, the number k in two extreme groups being chosen as near n/3 as possible, then using these two extreme groups to fit the slope. Furthermore in 1957, W. M. Gibson and G. H. Jowett [4] studied the problem of the optimum allocation of points in the three groups in Bartlett's method. They showed that the allocation of n/3 (approximately .33 of the data points) was optimum if the values of X come from a uniform distribution. For other distributions of the X data, other allocations are optimum.

The estimates of these two simple methods converge to the parameters when the number of data approach infinity; that is the estimates of both methods are consistent estimates of the parameters of the equation Y = a + bX. All these details are covered in Chapter II.

The primary result of this thesis is the extension of the techniques of both simple methods to multiple regression models. The equation studied in this thesis is

$$Z = a + bX + cY \tag{1-2}$$

where X and Y are independent variables and Z is the dependent variable. Since all three variables are unobservable we actually obtain x, y, and z, where $x = X + e_x$, $y = Y + e_y$, and $z = Z + e_z$, and e_x, e_y, and e_z are random errors with zero means and σ_x^2, σ_y^2, and σ_z^2 variances. By extending Wald's method and Bartlett's method consistent point estimates of the parameters a, b, c, σ_x^2, σ_y^2, and σ_z^2 are obtained. All these details will be covered in Chapter III.

Throughout this thesis we call the extension of Wald's method the two group method and the extension of Bartlett's method the three group method. In Chapter IV these methods are compared to least squares estimation in simulated problem for the model in equation (1-2). These studies indicate that all three methods give very similar results.

Chapter V contains the conclusions and some discussion of problems for future study.

CHAPTER II

FITTING A STRAIGHT LINE WHEN BOTH VARIABLES ARE SUBJECT TO ERROR

2.1 Introduction

In this chapter we shall consider the equation in which there exists a functional relationship between two mathematical variables that cannot be observed due to error of measurement. For example, suppose distance S and time T are related by the equation

$$S = a + bT$$

where a is the distance at time T = 0 and b is velocity. Now suppose that S and T are not observable but s and t can be observed where $s = S + e_s$ and $t = T + e_t$, and e_s and e_t are error of measurement. We can rewrite the equation as

$$s = a + bt + (e_s - be_t)$$

Also we may set the random term $(e_s - be_t)$ equal to e_e and write the equation as

$$s = a + bt + e_e$$

At first sight the above equation looks analogous to that usually solved by the least squares method or the maximum likelihood method. However here t is a random variable and not independent of the error term e_e, therefore this relation does not fit into the framework of either of the afore mentioned methods. If we add one more condition, such as the ratio of two variances σ_s^2 and σ_t^2 are known, then the maximum likelihood method are valid to solve this problem.

2.2 The General Assumptions

Suppose we obtain two sets of observations

$$x_1, x_2, \dots , x_n ; y_1, y_2, \dots , y_n$$

and we make the following assumptions:

(1) $x_i = X_i + e_{xi}$ and $y_i = Y_i + e_{yi}$, where X_i and Y_i are unobservable quantities and e_{xi} and e_{yi} are random variables for i = 1, ... , n. The pairs (x_1 , y_i) are observable.

(2) A single linear relation holds between X and Y, that is to say Y = a + bX.

(3) Each random variable $e_{x1}, \dots , e_{xn}$ has the same distribution with zero mean and σ_x^2 variance and the random variables are uncorrelated.

(4) Each random variable $e_{y1}, \dots , e_{yn}$ has the same distribution with zero mean and σ_y^2 variance and the random variables are uncorrelated.

(5) The random variables e_x and e_y are uncorrelated.

2.3 Wald's Method

With the above assumptions, Wald examined this problem in 1941 [2]. As the estimates of a, b, σ_x^2 , and σ_y^2 Wald used the following expressions:

(1)

$$\hat{b} = \frac{\dfrac{(y_{21} + \dots + y_{2m})}{m} - \dfrac{(y_{11} + \dots + y_{1m})}{m}}{\dfrac{(x_{21} + \dots + x_{2m})}{m} - \dfrac{(x_{11} + \dots + x_{1m})}{m}} \qquad (2\text{-}1)$$

where n is an even number, $m = n/2$, y_{2i} and y_{1j} correspond to x_{2i} and x_{1j}, x_{2i} is in the upper group (second group) with large values of x, x_{1j} is in the lower group (first group) with small values of x, $i = 1, \dots, m$, and $j = 1, \dots, m$.

(2)

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \qquad (2\text{-}2)$$

where $\bar{x}$ and $\bar{y}$ are the means of the variables x and y.

(3)

$$\hat{\sigma}_x^2 = (\dot{s}_x^2 - \frac{\dot{s}_{xy}}{\hat{b}}) \frac{n}{n - 1} \qquad (2\text{-}3)$$

$$\hat{\sigma}_y^2 = (\dot{s}_y^2 - \hat{b}\dot{s}_{xy}) \frac{n}{n - 1} \qquad (2\text{-}4)$$

where

$$\dot{s}_x^2 = \frac{\Sigma(x_i - \bar{x})^2}{n} , \quad \dot{s}_y^2 = \frac{\Sigma(y_i - \bar{y})^2}{n}$$

$$\dot{s}_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n}$$

$\hat{a}$, $\hat{b}$, $\hat{\sigma}_x^2$, and $\hat{\sigma}_y^2$ are the estimates of the parameters a, b, σ_x^2, and σ_y^2 respectively.

2.4 Bartlett's Method

(1)

$$\hat{b} = \frac{\bar{y}_3 - \bar{y}_1}{\bar{x}_3 - \bar{x}_1} \tag{2-5}$$

where

$$\bar{x}_3 = \frac{x_{31} + \ldots + x_{3k}}{k}, \quad \bar{x}_1 = \frac{x_{11} + \ldots + x_{1k}}{k}$$

and

$$\bar{y}_3 = \frac{y_{31} + \ldots + y_{3k}}{k}, \quad \bar{y}_1 = \frac{y_{11} + \ldots + y_{1k}}{k}$$

in which $k = n/3$, y_{3i} and y_{1j} correspond to x_{3i} and x_{1j}, x_{3i} is in the upper group (third group) with largest values of x, x_{1j} is in the lower group (first group) with smallest values of x, $i = 1, \ldots, k$, and $j = 1, \ldots, k$.

(2)

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \tag{2-6}$$

where $\bar{x}$ is the mean of the variable x, and $\bar{y}$ is the mean of the variable y.

$\hat{a}$ and $\hat{b}$ are the estimates of the parameters a and b.

2.5 The Optimum Choice of the Number of Points in the Three Groups

The number of data points in the three groups described above are equal; the ratio of points in the three groups is 1 : 1 : 1. Some people may question if it is necessary that the ratio has to be kept at the 1 : 1 : 1 standard for all situations. W. M. Gibson and G. H. Jowett [4], in 1957, examined several different distributions of X and found the optimum allocation of points in the three groups varied with the type of X-distribution.

The different distributions of the variable X that were examined by Gibson and Jowett are shown below:

(1)	Normal:	$f(X) = e^{-X^2/2}$	$-\infty < X < +\infty$
(2)	Uniform:	$f(X) = 1/2$	$-1 < X < +1$
(3)	Bell-shaped:	$f(X) = 1 - X^2$	$-1 < X < +1$
(4)	U-shaped:	$f(X) = 1/4 - X^4$	$-1 < X < +1$
(5)	J-shaped:	$f(X) = e^{-X}$	$-2 < X < +\infty$
(6)	Skew:	$f(X) = X^3 e^{-X/2}$	$0 < X < +\infty$

The optimum proportions of points in the three groups with different distributions of the variable X are given in Table 2-1.

Table 2-1

Optimum proportions of points in the three groups

distributions	proportions			approximate ratios		
	n_1	n_2	n_3	n_1	n_2	n_3
Normal	0.27	0.46	0.27	1	2	1
Uniform	0.33	0.33	0.33	1	1	1
Bell-shaped	0.31	0.38	0.31	3	4	3
U-shaped	0.39	0.22	0.39	2	1	2
J- shaped	0.45	0.40	0.15	3	3	1
Skew	0.36	0.45	0.19	4	5	2

2.6 Maximum Likelihood Method

We can use the maximum likelihood method [1] to obtain the estimates of the parameters of the equation Y = a + bX if we know the distributions of e_y and e_x are normal and have knowledge of the ratio $\lambda = \sigma_x^2/\sigma_y^2$. These estimates are consistent. This procedure is not considered because of the need to know $\lambda = \sigma_x^2/\sigma_y^2$.

CHAPTER III

FITTING THE EQUATION Z = a + bX + cY WHEN ALL VARIABLES ARE SUBJECT TO ERROR

3.1 Introduction

Up to this point we have considered the methods for fitting a straight line to the equation $Y = a + bX$. Usually more complex linear models are needed in practical situations. There are many problems in which a knowledge of more than one independent variable is necessary in order to obtain a better prediction of a particular response. The two simple methods given in Chapter II provide us with a procedure for extension to more complicated linear models. We shall apply both simple methods to the first-order linear model

$$Z = a + bX + cY \qquad (3\text{-}1)$$

3.1 The Formulation of the Problem

Let us begin with the precise formulation of the problem. We obtain three sets of observable variables

$$x_1, \ldots, x_n \; ; \; y_1, \ldots , y_n \; ; \; z_1, \ldots , z_n$$

and know that $x_i = X_i + e_{xi}$, $y_i = Y_i + e_{yi}$, and $z_i = Z_i + e_{zi}$, where X_i, Y_i, and Z_i are unobservable variables and e_{xi}, e_{yi}, and e_{zi} are unobservable random errors. We can observe the set $\{x_i, y_i, z_i\}$ for $i = 1, \ldots , n$ data points. Then we make the following assumptions:

(1) A linear relation holds among the unobservable variables X, Y, and Z; there is an equation $Z = a + bX + cY$.

(2) Each random variable $e_{x1}, \ldots, e_{xn}$ has the same distribution with zero mean and σ_x^2 variance and the variables are uncorrelated.

(3) Each random variable $e_{y1}, \ldots, e_{yn}$ has the same distribution with zero mean and σ_y^2 variance and the variables are uncorrelated.

(4) Each random variable $e_{z1}, \ldots, e_{zn}$ has the same distribution with zero mean and σ_z^2 variance and the variables are uncorrelated.

(5) The random variables e_x, e_y, and e_z are uncorrelated with one another.

(6) X and Y are uncorrelated with each other.

3.3 The Consistent Estimates of the Coefficients of the Equation by the Two Group Method

Using the above assumptions, the procedure of the two group method is as follows:

(1) Find the relationship between the first independent variable, say x, and the dependent variable z by Wald's method given in Section 2.3. The resulting equation is

$$\hat{z}_x = \hat{a}_{xz} + \hat{b}_{xz}x \qquad (3\text{-}2)$$

where

$$\hat{b}_{xz} = \frac{\frac{(z_{21} + \ldots + z_{2m})}{m} - \frac{(z_{11} + \ldots + z_{1m})}{m}}{\frac{(x_{21} + \ldots + x_{2m})}{m} - \frac{(x_{11} + \ldots + x_{1m})}{m}}$$

$$= \frac{\bar{z}_2 - \bar{z}_1}{\bar{x}_2 - \bar{x}_1} \qquad (3\text{-}2\text{-}1)$$

in which $m = n/2$, n is an even number, z_{2i} and z_{1j} correspond to x_{2i} and x_{1j}, x_{2i} is in the upper group (second group) with large values of x, x_{1j} is in the lower group (first group) with small values of x, $i = 1, \ldots, m$, and $j = 1, \ldots, m$,

and

$$\hat{a}_{xz} = \bar{z} - \hat{b}_{xz}\bar{x} \qquad (3\text{-}2\text{-}2)$$

in which $\bar{z}$ is the mean of the variable z and $\bar{x}$ is the mean of the variable x.

Finally, after $\hat{z}_x$ is obtained, calculate the residuals

$$z_i - \hat{z}_{xi}, \; i = 1, \ldots, n.$$

The fitted equation does not predict z exactly. However adding another variable y to the prediction equation might improve the prediction of z significantly. In order to accomplish this , we desire to relate the variable y to the unexplained variation in the variable z after the effect of the independent variable x has been removed from the variable z. However, if the variable x is in any way related to the variability shown in the variable y, we must correct for this first. Thus, what we need to do is to determine the relationship between the unknown variation in the variable z after the effect of the variable x has been removed and the remaining variation in the variable y after the effect of the variable x has been removed. Therefore in next step, we have y in term of x.

(2) Find the relationship between y and x. The resulting equation is

$$\hat{y}_x = \hat{a}_{xy} + \hat{b}_{xy}x \qquad (3\text{-}3)$$

where

$$\hat{b}_{xy} = \frac{\frac{(y_{21} + \ldots + y_{2m})}{m} - \frac{(y_{11} + \ldots + y_{1m})}{m}}{\frac{(x_{21} + \ldots + x_{2m})}{m} - \frac{(x_{11} + \ldots + x_{1m})}{m}}$$

$$= \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1} \qquad (3\text{-}3\text{-}1)$$

in which $m = n/2$, y_{2i} and y_{1j} correspond to x_{2i} and x_{1j}, x_{2i} is in the upper group with large values of x, x_{1j} is in the lower group with small values of x, $i = 1, \ldots, m$, and $j = 1, \ldots, m$, and

$$\hat{a}_{xy} = \bar{y} - \hat{b}_{xy}\bar{x} \qquad (3\text{-}3\text{-}2)$$

in which $\bar{y}$ is the mean of the variable y and $\bar{x}$ is the mean of the variable x.

Finally, after $\hat{y}_x$ is obtained, calculate the residuals

$$y_i - \hat{y}_{xi},\ i = 1, \ldots , n.$$

(3) Find the relationship between $(z - \hat{z}_x)$ and $(y - \hat{y}_x)$ by Wald's method. The resulting equation is

$$(z - \hat{z}_x) = \hat{b}_{yz}(y - \hat{y}_x) \qquad (3\text{-}4)$$

where

$$\hat{b}_{yz} = \frac{\dfrac{\Sigma(z - \hat{z}_x)_{2i}}{m} - \dfrac{\Sigma(z - \hat{z}_x)_{1j}}{m}}{\dfrac{\Sigma(y - \hat{y}_x)_{2i}}{m} - \dfrac{\Sigma(y - \hat{y}_x)_{1j}}{m}} \qquad (3\text{-}4\text{-}1)$$

in which $(z - \hat{z}_x)_{2i}$ and $(z - \hat{z}_x)_{1j}$ correspond to $(y - \hat{y}_x)_{2i}$ and $(y - \hat{y}_x)_{1j}$, $(y - \hat{y}_x)_{2i}$ is in the upper group, $(y - \hat{y}_x)_{1j}$ is in the lower group, $i = 1, \ldots , m$, and $j = 1, \ldots , m$.

Note that no $\hat{a}_{yz}$ term in equation (3-4) is required since we use two sets of the residuals whose average values are zero.

(4) Within the parenthesis of equation (3-4) we substitute $(\hat{a}_{xz} + \hat{b}_{xz}x)$ for $\hat{z}_x$ and $(\hat{a}_{xy} + \hat{b}_{xy}x)$ for $\hat{y}_x$. The

resulting equation is

$$(z - (\hat{a}_{xz} + \hat{b}_{xz}x)) = \hat{b}_{yz}(y - (\hat{a}_{xy} + \hat{b}_{xy}x))$$

After shifting $(\hat{a}_{xz} + \hat{b}_{xz}x)$ to the right side and rearranging the above equation, we have

$$z = (\hat{a}_{xz} - \hat{b}_{yz}\hat{a}_{xy}) + (\hat{b}_{xz} - \hat{b}_{yz}\hat{b}_{xy})x + \hat{b}_{yz}y$$

Simplifing the above equation, we have the prediction equation

$$z = \hat{a} + \hat{b}x + \hat{c}y \tag{3-5}$$

where

$$\hat{a} = \hat{a}_{xz} - \hat{b}_{yz}\hat{a}_{xy} \tag{3-5-1}$$

$$\hat{b} = \hat{b}_{xz} - \hat{b}_{yz}\hat{b}_{xy} \tag{3-5-2}$$

$$\hat{c} = \hat{b}_{yz} \tag{3=5-3}$$

Next we shall prove that $\hat{a}$, $\hat{b}$, and $\hat{c}$ are the consistent estimates of a, b, and c; that is $\hat{a}$, $\hat{b}$, and $\hat{c}$ converge to a, b, and c when $n \rightarrow \infty$, First consider $\hat{a}_{xz}$ and $\hat{b}_{xz}$ when $n \rightarrow \infty$. If all assumptions in Section 3.2 are true, then equation (3-1) will become

$$z = a + bx - be_x + cy - ce_y + e_z$$

When $n \rightarrow \infty$, the mean of the variable z will be

$$\lim_{n\to\infty} \bar{z} = \frac{\Sigma z_i}{n} = \frac{\Sigma(a + bx_i - be_{xi} + cy_i - ce_{yi} + e_{zi})}{n}$$

$$= a + b\frac{\Sigma x_i}{n} + c\frac{\Sigma y_i}{n} - b\frac{\Sigma e_{xi}}{n} - c\frac{\Sigma e_{yi}}{n} + \frac{\Sigma e_{zi}}{n}$$

$$= a + b\bar{x} + c\bar{y} \qquad (3\text{-}6)$$

because $\lim_{n\to\infty}\frac{\Sigma e_{xi}}{n}$, $\lim_{n\to\infty}\frac{\Sigma e_{yi}}{n}$, and $\lim_{n\to\infty}\frac{\Sigma e_{zi}}{n}$ are all zero. From equation (3-2-1) we have

$$\hat{b}_{xz} = \frac{\bar{z}_2 - \bar{z}_1}{\bar{x}_2 - \bar{x}_1}$$

Substituting $(a + b\bar{x} + c\bar{y})$ from equation (3-6) for $\bar{z}_2$ and $\bar{z}_1$ in the numerator of the above equation, then we have

$$\lim_{n\to\infty} \hat{b}_{xz} = \frac{(a + b\bar{x}_2 + c\bar{y}_2) - (a + b\bar{x}_1 + c\bar{y}_1)}{\bar{x}_2 - \bar{x}_1}$$

$$= b\frac{\bar{x}_2 - \bar{x}_1}{\bar{x}_2 - \bar{x}_1} + c\frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1}$$

According to the assumption (6) X and Y are uncorrelated with each other, therefore the value of the variable x which is divided into an upper group and a lower group does not affect the value of the variable y. Thus when n becomes large, the means of the variable y in the two groups are equal, that is

$\bar{y}_2 = \bar{y}_1$. Hence we have

$$\lim_{n\to\infty} \hat{b}_{xz} = b \frac{\bar{x}_2 - \bar{x}_1}{\bar{x}_2 - \bar{x}_1} = b \qquad (3-7)$$

From equation (3-2-2) we have

$$\hat{a}_{xz} = \bar{z} - \hat{b}_{xz}\bar{x}$$

As n becomes large, we substitute $(a + b\bar{x} + c\bar{y})$ for $\bar{z}$ and equation (3-7) for $\hat{b}_{xz}$, then we have

$$\begin{aligned}\lim_{n\to\infty} \hat{a}_{xz} &= a + b\bar{x} + c\bar{y} - b\bar{x} \\ &= a + c\bar{y} \qquad (3-8)\end{aligned}$$

Now consider $\hat{a}_{xy}$ and $\hat{b}_{xy}$ when $n \to \infty$. From equation (3-3-1) we have

$$\hat{b}_{xy} = \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1}$$

Again when n becomes very large, $\bar{y}_2 = \bar{y}_1$ thus we have

$$\lim_{n\to\infty} \hat{b}_{xy} = 0 \qquad (3-9)$$

From equation (3-3-2) we have

$$\hat{a}_{xy} = \bar{y} - \hat{b}_{xy}\bar{x}$$

When n becomes large, substituting equation (3-9) into the second term in the right side of the equation, we have

$$\lim_{n\to\infty} \hat{a}_{xy} = \bar{y} - 0\,\bar{x}$$

$$= \bar{y} \qquad (3\text{-}10)$$

Finally, obtaining $\lim_{n\to\infty} \hat{a}_{xz}$, $\lim_{n\to\infty} \hat{b}_{xz}$, $\lim_{n\to\infty} \hat{a}_{xy}$, and $\lim_{n\to\infty} \hat{b}_{xy}$, we substitute $\lim_{n\to\infty} \hat{a}_{xy}$ for the first term and $\lim_{n\to\infty} \hat{b}_{xy}$ for the second term of equation (3-3), then we have

$$\lim_{n\to\infty} \hat{y}_x = \bar{y} - 0\,\bar{x}$$

$$= \bar{y} \qquad (3\text{-}11)$$

We substitute $\lim_{n\to\infty} \hat{a}_{xz}$ for the first term and $\lim_{n\to\infty} \hat{b}_{xz}$ for the second term of equation (3-2), then we have

$$\lim_{n\to\infty} \hat{z}_x = (a + c\bar{y}) + bx$$

$$= a + c\bar{y} + bx \qquad (3\text{-}12)$$

After having $\lim_{n\to\infty} \hat{z}_x$ and $\lim_{n\to\infty} \hat{y}_x$, we can obtain $\lim_{n\to\infty} \hat{b}_{yz}$. From equation (3-4-1) we have

$$\hat{b}_{yz} = \frac{\dfrac{\Sigma(z - \hat{z}_x)_{2i}}{m} - \dfrac{\Sigma(z - \hat{z}_x)_{1j}}{m}}{\dfrac{\Sigma(y - \hat{y}_x)_{2i}}{m} - \dfrac{\Sigma(y - \hat{y}_x)_{1j}}{m}}$$

As n becomes large, substituting $(a + c\bar{y} + bx)$ from equation

(3-12) for $\hat{z}_x$, and equation (3-11) for $\hat{y}_x$, we have

$$\lim_{n\to\infty} \hat{b}_{yz} = \frac{\dfrac{\Sigma(z_{2i} - (a+c\bar{y}_2+bx_{2i}))}{m} - \dfrac{\Sigma(z_{1j} - (a+c\bar{y}_1+bx_{1j}))}{m}}{\dfrac{\Sigma(y_{2i} - \bar{y}_2)}{m} - \dfrac{\Sigma(y_{1j} - \bar{y}_1)}{m}}$$

$$= \frac{(\dfrac{\Sigma z_{2i}}{m} - \dfrac{\Sigma z_{1j}}{m}) - b(\dfrac{\Sigma x_{2i}}{m} - \dfrac{\Sigma x_{1j}}{m})}{\dfrac{\Sigma y_{2i}}{m} - \dfrac{\Sigma y_{1j}}{m}}$$

$$= \frac{(\bar{z}_2 - \bar{z}_1) - b(\bar{x}_2 - \bar{x}_1)}{\bar{y}_2 - \bar{y}_1}$$

and substituting $(a + b\bar{x} + c\bar{y})$ from equation (3-6) for $\bar{z}_2$ and $\bar{z}_1$, we have

$$\lim_{n\to\infty} \hat{b}_{yz} = \frac{(a + b\bar{x}_2 + c\bar{y}_2) - (a + b\bar{x}_1 + c\bar{y}_1) - b(\bar{x}_2 - \bar{x}_1)}{\bar{y}_2 - \bar{y}_1}$$

$$= \frac{c(\bar{y}_2 - \bar{y}_1)}{\bar{y}_2 - \bar{y}_1}$$

$$= c \qquad (3\text{-}13)$$

From the aforementioned discussion, we know when n becomes large, $\hat{a}_{xz}$ converges to $(a + c\bar{y})$, $\hat{b}_{xz}$ to b, $\hat{a}_{xy}$ to $\bar{y}$, $\hat{b}_{xy}$ to 0, and $\hat{b}_{yz}$ to c. Hence substituting all these values into equation (3-5-1), equation (3-5-2), and equation(3-5-3), we have

$$\lim_{n\to\infty} \hat{a} = (\lim_{n\to\infty} \hat{a}_{xz}) - (\lim_{n\to\infty} \hat{b}_{yz})(\lim_{n\to\infty} \hat{a}_{xy})$$

$$= (a + c\bar{y}) - c\bar{y}$$

$$= a$$

$$\lim_{n\to\infty} \hat{b} = (\lim_{n\to\infty} \hat{b}_{xz}) - (\lim_{n\to\infty} \hat{b}_{yz})(\lim_{n\to\infty} \hat{b}_{xy})$$

$$= b - c\ 0$$

$$= b$$

$$\lim_{n\to\infty} \hat{c} = \lim_{n\to\infty} \hat{b}_{yz}$$

$$= c$$

Thus from the above expressions we can be assured that $\hat{a}$, $\hat{b}$, and $\hat{c}$ are consistent estimates of the coefficients of the equation $Z = a + bX + cY$.

3.4 The Consistnet Estimates of the Coefficients of the Equation by the Three Group Method

Bartlett's method is more effective in estimating the parameters of the equation $Y = a + bX$ than Wald's therefore it may be benefical to extend his method to the equation $Z = a + bX + cY$. Comparing the two group method and the three group method, we see these two simple methods are very similar in procedure. Therefore we will briefly discuss the procedure of the three group method in the following sections. To illustrate the procedure we consider the case where X and Y both have uniform distribution.

The procedure of the three group method is as follows:

(i) Find the relationship between the first independent variable, say x, and the dependent variable z by Bartlett's method given in Section 2.4. The resulting equation is

$$\hat{z}_x = \hat{a}_{xz} + \hat{b}_{xz}x \tag{3-14}$$

where

$$\hat{b}_{xz} = \frac{\dfrac{z_{31} + \dots + z_{3k}}{k} - \dfrac{z_{11} + \dots + z_{1k}}{k}}{\dfrac{x_{31} + \dots + x_{3k}}{k} - \dfrac{x_{11} + \dots + x_{1k}}{k}}$$

$$= \frac{\bar{z}_3 - \bar{z}_1}{\bar{x}_3 - \bar{x}_1} \tag{3-14-1}$$

in which $k = n/3$, z_{3i} and z_{1j} correspond to x_{3i} and x_{1j}, x_{3i} is in the upper group (third group) with largest values of x, x_{1j} is in the lower group (first group) with smallest values of x, $i = 1, \dots,$ k, and $j = 1, \dots, k$,

and

$$\hat{a}_{xz} = \bar{z} - \hat{b}_{xz}\bar{x} \qquad (3\text{-}14\text{-}2)$$

in which $\bar{z}$ is the mean of the variable z and $\bar{x}$ is the mean of the variable x.

After $\hat{z}_x$ is obtained, calculate the residuals

$$z_i - \hat{z}_{xi},\ i = 1, \dots , n.$$

(2) Find the relationship between the variable y and the variable x. The resulting equation is

$$\hat{y}_x = \hat{a}_{xy} + \hat{b}_{xy}x \qquad (3\text{-}15)$$

where

$$\hat{b}_{xy} = \frac{\dfrac{y_{31} + \dots + y_{3k}}{k} - \dfrac{y_{11} + \dots + y_{1k}}{k}}{\dfrac{x_{31} + \dots + x_{3k}}{k} - \dfrac{x_{11} + \dots + x_{1k}}{k}}$$

$$= \frac{\bar{y}_3 - \bar{y}_1}{\bar{x}_3 - \bar{x}_1} \qquad (3\text{-}15\text{-}1)$$

in which $k = n/3$, y_{3i} and y_{1j} correspond to x_{3i} and x_{1j}, x_{3i} is in the upper group with largest values of x, x_{1j} is in the lower group with smallest values of x, $i = 1, \dots , k$, and $j = 1, \dots , k$, and

$$\hat{a}_{xy} = \bar{y} - \hat{b}_{xy}\bar{x} \qquad (3\text{-}15\text{-}2)$$

Finally calculate the residuals

$$y_i - \hat{y}_{xi}, \; i = 1, \ldots , n.$$

(3) Find the relationship between $(z - \hat{z}_x)$ and $(y - \hat{y}_x)$. The resulting equation is

$$(z - \hat{z}_x) = \hat{b}_{yz}(y - \hat{y}_x) \qquad (3\text{-}16).$$

where

$$\hat{b}_{yz} = \frac{\dfrac{\Sigma(z - \hat{z}_x)_{3i}}{k} - \dfrac{\Sigma(z - \hat{z}_x)_{1j}}{k}}{\dfrac{\Sigma(y - \hat{y}_x)_{3i}}{k} - \dfrac{\Sigma(y - \hat{y}_x)_{1j}}{k}}$$

in which $(z - \hat{z}_x)_{3i}$ and $(z - \hat{z}_x)_{1j}$ correspond to $(y - \hat{y}_x)_{3i}$ and $(y - \hat{y}_x)_{1j}$, $(y - \hat{y}_x)_{3i}$ is in the upper group with largest values of $(y - \hat{y}_x)$, $(y - \hat{y}_x)_{1j}$ is in the lower group with smallest values of $(y - \hat{y}_x)$, and $k = n/3$.

(4) Within the parenthesis of equation (3-16), we substitute $(\hat{a}_{xz} + \hat{b}_{xz}x)$ from equation (3-14) for $\hat{z}_x$ and $(\hat{a}_{xy} + \hat{b}_{xy}x)$ from equation (3-15) for $\hat{y}_x$ in equation (3-16). After simplification the result becomes similar to (3-5), that is

$$z = \hat{a} + \hat{b}x + \hat{c}y \qquad (3\text{-}17)$$

where

$$\hat{a} = \hat{a}_{xz} - \hat{b}_{yz}\hat{a}_{xy} \qquad (3\text{-}17\text{-}1)$$

$$\hat{b} = \hat{b}_{xz} - \hat{b}_{yz}\hat{b}_{xy} \qquad (3\text{-}17\text{-}2)$$

$$\hat{c} = \hat{b}_{yz} \qquad (3{=}17\text{-}3)$$

As shown in the preceding section, we also can prove $\hat{a}$, $\hat{b}$, and $\hat{c}$ are the consistent estimates of the coefficients of the equation Z = a + bX + cY.

3.5 Example

We will look at an example solved by the two group method and the three group method. Now suppose the distance a particle travels from a given reference point is given theoretically by the curve

$$Z = a + bX + cY$$

where Z is the distance, X is the time a particle moves, and Y is the temperature of the medium through which the particle moves. These three variables are all measured with error. We observe x, y, and z where $x = X + e_x$, $y = Y + e_y$, and $z = Z + e_z$. $\ddot{e}_x$, e_y, and e_z are normal random errors with zero means and σ_x^2, σ_y^2, and σ_z^2 variances. The data is given in Table 3-1.

Table 3-1

16 observations

of x, y, and z

observation number	x	y	z
1	10.043	9.318	49.222
2	11.723	10.362	55.480
3	8.580*	9.740*	42.004*
4	10.519	5.722	39.126
5	6.061*	9.792*	44.543*
6	3.880*	10.638*	45.688*
7	5.615*	2.229*	22.833*
8	3.426*	6.454*	30.729*
9	5.789*	11.538*	51.718*
10	9.708	9.743	51.203
11	14.942	6.139	48.328
12	10.514	-0.800	19.831
13	11.348	4.190	36.634
14	7.876*	11.485*	52.518*
15	9.597	3.272	33.330
16	5.535*	2.859*	23.791*

data with * is in the lower group

The solution of two group method:

(1) Find the relationship between z and x by the two group method. The result is

$$\hat{z}_x = \hat{a}_{xz} + \hat{b}_{xz}x$$

$$= 37.17 + 0.39x$$

Then calculate the residuals, $z_i - \hat{z}_{xi}$, $i = 1, \ldots, 16$. These residuals are shown in Table 3-2.

Table 3-2

Residuals: $z_i - \hat{z}_{xi}$

observation number	x	z	$z_i - \hat{z}_{xi}$
1	10.043	49.222	12.244
2	11.723	55.480	13.649
3	8.580	42.004	3.423
4	10.519	39.126	-2.190
5	6.061	44.543	4.983
6	3.880	45.688	6.988
7	5.615	22.833	-16.552
8	3.426	30.729	-7.792
9	5.789	51.718	12.266
10	9.708	51.203	9.207
11	14.942	48.328	5.268
12	10.514	19.831	-21.483
13	11.348	36.634	-4.973
14	7.876	52.518	12.244
15	9.597	33.330	-7.623
16	5.535	23.791	-15.560

(2) Find the relationship between y and x. The result is

$$\hat{y}_x = \hat{a}_{xy} + \hat{b}_{xy}x$$

$$= 9.96 - 0.359x$$

Then calculate the residuals, $y_i - \hat{y}_{xi}$, $i = 1, \ldots, 16$. These residuals are shown in Table 3-3.

Table 3-3

Residuals: $y_i - \hat{y}_{xi}$

observation number	x	y	$y_i - \hat{y}_{xi}$
1	10.043	9.318	3.094
2	11.723	10.362	4.618
3	8.580	9.740	1.161
4	10.519	5.722	-0.470
5	6.061	9.792	1.903
6	3.880	10.638	2.065
7	5.615	2.229	-5.721
8	3.426	6.454	-2.281
9	5.789	11.538	3.651
10	9.708	9.743	3.261
11	14.942	6.139	1.534
12	10.514	-0.800	-6.993
13	11.348	4.190	-1.640
14	7.876	11.485	4.346
15	9.597	3.272	-3.250
16	5.535	2.859	-5.121

(3) Find the relationship between two sets of the residuals $z_i - \hat{z}_{xi}$ and $y_i - \hat{y}_{xi}$. These two sets of the residuals are shown in Table 3-4. From Table 3-4, using the two group method, the result is

$$(z - \hat{z}_x) = \hat{b}_{yz}(y - \hat{y}_x)$$

$$= 2.97(y - \hat{y}_x)$$

Table 3-4

Residuals: $z_i - \hat{z}_{xi}$ and $y_i - \hat{y}_{xi}$

observation number	$y_i - \hat{y}_{xi}$	$z_i - \hat{z}_{xi}$
1	3.095	12.244
2	4.618	18.698
3	1.161*	3.423*
4	-0.470*	-2.190*
5	1.903	4.983
6	2.065	6.988
7	-5.721*	-16.552*
8	-2.281*	-7.792*
9	3.651	12.266
10	3.261	9.207
11	1.534	5.268
12	-6.993*	-21.483*
13	-1.640*	-4.973*
14	4.346	12.244
15	-3.250*	-7.792*
16	-5.121*	-15.500*

(4) Within the parenthesis of the equation, substituting (37.17 + 0.39x) for $\hat{z}_x$ and (9.96 - 0.359x) for $\hat{y}_x$, the result is

$$z - (37.17 + 0.39x) = 2.97\ (y - (9.96 - 0.359x))$$

$$z = 7.42 + 1.46x + 2.97\ y$$

The above solution is the prediction equation solved by the two group method.

If we use the three group method to solve this problem, we get another solution

$$z = 7.863 + 1.429x + 2.964y$$

3.6 The Consistent Estimates of σ_x^2, σ_y^2, and σ_z^2

Let us introduce the following notations:

$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n - 1}} = \text{the sample standard deviation of } x$$

$$s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n - 1}} = \text{the sample standard deviation of } y$$

$$s_z = \sqrt{\frac{\Sigma(z_i - \bar{z})^2}{n - 1}} = \text{the sample standard deviation of } z$$

$$s_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n - 1} = \text{the sample covariance between } x \text{ and } y$$

$$s_{xz} = \frac{\Sigma(x_i - \bar{x})(z_i - \bar{z})}{n - 1} = \text{the sample covariance between } x \text{ and } z$$

$$s_{yz} = \frac{\Sigma(y_i - \bar{y})(z_i - \bar{z})}{n - 1} = \text{the sample covariance between } y \text{ and } z$$

S_X, S_Y, S_Z, S_{XY}, S_{XZ}, and S_{YZ} denote the same expressions for the unobservable values of X, Y, and Z.

Now we have equations of expection, which are proved in Appendix A where all expections are over the distributions of the measurement errors, e_x, e_y, and e_z, as follows:

$$E(s_x^2) = S_X^2 + \sigma_x^2 \qquad (3\text{-}18\text{-}1)$$

$$E(s_y^2) = S_Y^2 + \sigma_y^2 \qquad (3\text{-}18\text{-}2)$$

$$E(S_z^2) = S_Z^2 + \sigma_z^2 \qquad (3\text{-}18\text{-}3)$$

$$E(S_{xy}) = S_{XY} \qquad (3\text{-}18\text{-}4)$$

$$E(S_{xz}) = S_{XZ} \qquad (3\text{-}18\text{-}5)$$

$$E(S_{zy}) = S_{YZ} \qquad (3\text{-}18\text{-}6)$$

Since Z = a + bX + cY, we have the following equations, which also are proved in Appendix A,

$$S_Z^2 = b^2S_X^2 + c^2S_Y^2 + 2bcS_{XY} \qquad (3\text{-}18\text{-}7)$$

$$S_{XZ} = bS_X^2 + cS_{XY} \qquad (3\text{-}18\text{-}8)$$

$$S_{YZ} = cS_Y^2 + bS_{XY} \qquad (3\text{-}18\text{-}9)$$

We solve equation (3-18-7), equation (3-18-8), and equation (3-18-9) and have S_Z^2 in terms of S_{XZ} and S_{YZ}, S_X^2 in terms of S_{XZ} and S_{XY}, and S_Y^2 in terms of S_{YZ} and S_{XY} as follows:

$$S_Z^2 = bS_{XZ} + cS_{YZ} \qquad (3\text{-}18\text{-}10)$$

$$S_X^2 = \frac{1}{b}S_{XZ} - \frac{c}{b}S_{XY} \qquad (3\text{-}18\text{-}11)$$

$$S_Y^2 = \frac{1}{c}S_{YZ} - \frac{b}{c}S_{XY} \qquad (3\text{-}18\text{-}12)$$

Substituting $E(S_{xy})$ from equation (3-18-4) for S_{XY}, $E(S_{xz})$ from equation (3-18-5) for S_{XZ}, and $E(S_{yz})$ from equation (3-18-6) for S_{YZ} in the above equations, we have

$$S_Z^2 = bE(S_{xz}) + cE(S_{yz}) \tag{3-18-13}$$

$$S_X^2 = \frac{1}{b}E(S_{xz}) - \frac{c}{b}E(S_{xy}) \tag{3-18-14}$$

$$S_Y^2 = \frac{1}{c}E(S_{yz}) - \frac{b}{c}E(S_{xy}) \tag{3-18-15}$$

Substituting the right side of equation (3-18-13) for S_Z^2 in equation (3-18-3), the right side of equation (3-18-14) for S_X^2 in equation (3-18-1), and the right side of equation (3-18-15) for S_Y^2 in equation (3-18-2) and rearranging these three equations, we have

$$\sigma_x^2 = E(S_x^2) - \frac{1}{b}E(S_{xz}) + \frac{c}{b}E(S_{xy}) \tag{3-18-16}$$

$$\sigma_y^2 = E(S_y^2) - \frac{1}{c}E(S_{yz}) + \frac{b}{c}E(S_{xy}) \tag{3-18-17}$$

$$\sigma_z^2 = E(S_z^2) - bE(S_{xz}) - cE(S_{yz}) \tag{3-18-18}$$

Since when n becomes very large, S_x^2, S_y^2, S_z^2, S_{xz}, S_{xy}, and S_{yz} converge toward their expected values and $\hat{a}$, $\hat{b}$, and $\hat{c}$ converge toward a, b, and c, the expressions

$$\hat{\sigma}_x^2 = (S_x^2 - \frac{1}{\hat{b}}S_{xz} + \frac{\hat{c}}{\hat{b}}S_{xy})$$

$$\hat{\sigma}_y^2 = (S_y^2 - \frac{1}{\hat{c}}S_{yz} + \frac{\hat{b}}{\hat{c}}S_{xy})$$

$$\hat{\sigma}_z^2 = (S_z^2 - \hat{b}S_{xz} - \hat{c}S_{yz})$$

are the consistent estimates of σ_x^2, σ_y^2, and σ_z^2 respectively.

3.7 Fitting the Equation $W = a + bX + cY + dZ$ with All Variables Subject to Error

Consider briefly the two simple methods when applied to the equation

$$W = a + bX + cY + dZ$$

Now we obtain four sets of observations

$$x_1, \ldots, x_n ; y_1, \ldots, y_n ; z_1, \ldots, z_n ;$$

$$w_1, \ldots, w_n$$

where $x = X + e_x$, $y = Y + e_y$, $z = Z + e_z$, and $w = W + e_w$. X, Y, Z, and W are unobservable variables and e_x, e_y, e_z, and e_w are unobservable random errors. The assumptions listed in Section 3.2 are extended as follows:

(1) A linear relation holds among the unobservable values X, Y, Z, and W; there is an equation $W = a + bX + cY + dZ$.

(2) Assumptions (2), (3), and (4) still hold.

(3) Each random variable $e_{w1}, \ldots, e_{wn}$ has the same distribution with zero mean and σ_ω^2 variance and the variables are uncorrelated.

(4) The random variables e_x, e_y, e_z, and e_w are uncorrelated.

(5) X, Y, and Z are uncorrelated with one another.

Suppose the above assumptions are true and the observations of the $\{x, y, z, w\}$ data set are obtained, the procedure for fitting the equation $W = a + bX + cY + dZ$ is as follows:

(1) Find the relationship among the first two variables, say x and y, and the dependent variable w by the two group method (or the three group method). The resulting equation is

$$\hat{w}_1 = \hat{a}_1 + \hat{b}_1 x + \hat{c}_1 y \qquad (3\text{-}19)$$

After $\hat{w}_1$ is obtained, calculate the residuals

$$w_1 - \hat{w}_{1i}, \; i = 1, \ldots , n.$$

As discussed in Section 3.3 the above fitted equation does not predict w exactly. Adding the other variable z to the prediction equation may improve the prediction of w. Thus, what we need to do is to determine the relationship between the unknown variation in the variable w after the effect of the variables x and y has been removed from w and the remaining variation in the variable z after the effect of the variables x and y has been removed from z. Therefore, in the next step we have z in terms of x and y.

(2) Find the relationship among the first two variables, x and y, and the variable z by the two group method (or the three group method). The resulting equation is

$$\hat{z}_2 = \hat{a}_2 + \hat{b}_2 x + \hat{c}_2 y \qquad (3\text{-}20)$$

After $\hat{z}_2$ is obtained, calculate $z_i - \hat{z}_{2i}$, $i = 1, \ldots, n$.

(3) Find the relationship between $(w - \hat{w}_1)$ and $(z - \hat{z}_2)$ by the two group method (or the three group method). The resulting equation is

$$(w - \hat{w}_1) = \hat{b}_3(z - \hat{z}_2) \qquad (3\text{-}21)$$

(4) Within the parenthesis of equation (3-21) substituting $(\hat{a}_1 + \hat{b}_1 x + \hat{c}_1 y)$ from equation (3-19) for $\hat{w}_1$ and $(\hat{a}_2 + \hat{b}_2 x + \hat{c}_2 y)$ from equation (3-20) for $\hat{z}_2$, the resulting equation is

$$(w - (\hat{a}_1 + \hat{b}_1 x + \hat{c}_1 y)) = \hat{b}_3(z - (\hat{a}_2 + \hat{b}_2 x + \hat{c}_2 y))$$

After shifting $(\hat{a}_1 + \hat{b}_1 x + \hat{c}_1 y)$ to the right side and rearranging the above equation, it is

$$w = (\hat{a}_1 - \hat{a}_2\hat{b}_3) + (\hat{b}_1 - \hat{b}_2\hat{b}_3)x + (\hat{c}_1 - \hat{c}_2\hat{b}_3)y + \hat{b}_3 z$$

After simplification the above prediction equation becomes

$$w = \hat{a} + \hat{b}x + \hat{c}y + \hat{d}z$$

where

$$\hat{a} = \hat{a}_1 - \hat{a}_2\hat{b}_3$$

$$\hat{b} = \hat{b}_1 - \hat{b}_2\hat{b}_3$$

$$\hat{c} = \hat{c}_1 - \hat{c}_2\hat{b}_3$$

$$\hat{d} = \hat{b}_3$$

$\hat{a}$, $\hat{b}$, $\hat{c}$, and $\hat{d}$ are the estimates of the coefficients of the equation $W = a + bX + cY + dZ$.

CHAPTER IV

SIMULATION

A FORTRAN IV program has been written to fit the linear equation Z = a + bX + cY by the two group method, the three group method, and the least squares method. The observable values of the variables X, Y, and Z are generated in program and these observable values come from a linear equation Z = a + bX + cY. where a, b, and c are selected and known values.

In our program XX, YY, and ZZ denote unobservable variables, X, Y, and Z denote observable variables, and EX, EY, and EZ are for random error terms. First we generate XX and YY from uniform distribution within their region. This makes Bartlett's method consistent with the optimum proportion 1 : 1 : 1 given by Gibson and Jowett. The unobservable dependent variable is ZZ = a + bXX + cYY. Next we generate normal random values with zero means and σ_x^2 , σ_y^2 , and σ_z^2 variances for the error terms EX, EY, and EZ. The observable values in our program are X = XX + EX, Y = YY + EY, and Z = ZZ + EZ. We can use this procedure to generate data consistent with the assumptions of our model. The program is shown in Appendix B.

In the example shown below we initially give the following input data to the program,

the region of XX is (17 to 1)

the region of YY is (20 to 7)

the standard deviation of EX, EY and Ez are 0.1

a = 4, b = -1, and c = 2

the number of observations is 10

Then we obtain 10 observations of X, Y, Z, XX, YY, and ZZ as shown in Table 4-1

Table 4-1

X	Y	Z	XX	YY	ZZ
15.63	17.54	23.13	15.81	17.44	23.07
15.08	7.04	3.14	15.03	7.13	3.23
15.80	9.69	7.71	15.83	9.79	7.74
12.64	11.84	15.19	12.72	11.84	14.97
12.17	13.26	18.50	12.19	13.33	18.48
3.40	12.67	26.42	3.24	12.83	26.42
12.91	8.04	7.07	12.87	8.01	7.16
14.31	9.31	8.35	14.40	9.28	8.16
3.45	7.67	16.09	3.21	7.64	16.07
14.86	13.34	16.01	14.80	13.38	15.95

The subroutine REARR is used to rearrange an array in increasing order. The CALL REARR(F1,F2,F3,F4,F5,F6,NO) statement will rearrange according to the specified variable, F1, meanwhile F2, F3, F4, F5, and F6 will be also rearranged according to the new order of F1 values. In our program X is F1. NO is the number of observations and F2, F3, F4, F5, and F6 all correspond to F1. After rearrangement it is easy to divide the set into either two or three groups. Under the new order, the values of X, Y, Z, XX, YY, and ZZ are shown in

Table 4-2.

The subroutine TAT calculate the estimates using the data as sorted. The CALL TAT(X,Y,Z,XAVE,YAVE,ZAVE,NO,A,B,C,L) statement will make the points estimates of the coefficients of the equation by the two group method and the three group method. XAVE, YAVE, and ZAVE are the average of X, Y, and Z. NO is the number of observations, L = 1 is for the two group method, and L = 2 is for the three group method. A, B, and C given by the subroutine are the estimates of a, b, and c. The rest of the arguments are furnished by the CALL statement.

Table 4-2

X	Y	Z	XX	YY	ZZ
3.40*	12.67*	26.42*	3.24	12.83	26.42
3.45*	7.67*	16.09*	3.21	7.64	16.07
12.17*	13.26*	18.50*	12.19	13.33	18.48
12.64	11.84	15.19	12.72	11.84	14.97
12.91	8.04	7.07	12.87	8.01	7.16
14.31	9.31	8.35	14.40	9.28	8.16
14.86	13.34	16.01	14.80	13.38	15.95
15.08#	7.04#	3.14#	15.03	7.13	3.23
15.63#	17.54#	23.13#	15.81	17.44	23.07
15.80#	9.69#	7.71#	15.83	9.79	7.74

data with # is in the third group
data with * is in the first group

We also determine the estimates by the least squares method using the subroutine LSM. The CALL LSM(X,Y,Z,XAVE,YAVE,ZAVE, A,B,C,NO) statement will make the point estimates of the coefficients by the least squares method. All the arguments in the statement CALL LSM are the same as those in CALL TAT, except without L. All SUBROUTINE programs are shown in Appendix B. The results for these 10 observations shown in Table 4-1 are:

two group method:	a = 4.32	b = -1.02	c = 2.00
three group method:	a = 4.44	b = -1.03	c = 2.00
least squares method:	a = 4.55	b = -1.02	c = 1.98

A series of simulated example problems have been run similar to the example above. The results have been combined in Appendix C. The method of calculating the values in Appendix C is as follows:

In the first row, the set (A = 1.86, B = 4.06, C = 2.03) is the average estimate of three runs by the two group method with 10 data points and $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = 0.1$. The second set (A = 2.35, B = 4.03, C = 2.01) is the average estimate of three runs by the three group method using the same three sets of 10 data points. The third set (A = 2.30, B = 4.03, C = 2.01) is the average estimate of three runs by the least squares method with the same data sets, and the other sets are similarly the average estimates of three runs with 10 data points but with different σ_x^2, σ_y^2, and σ_z^2. In the output. NO is the number of observations, STD are standard deviations σ_x, σ_y, and σ_z. We let $\sigma_x = \sigma_y = \sigma_z$ throughout our problems.

In Appendix C, the problem are

(1) fit the equation Z = 3 + 4X + 2Y with 10 data points

(2) fit the equation $Z = -3 + 3X - 4Y$ with 20 data points

(3) fit the equation $Z = 5 - 4X + 2Y$ with 30 data points

(4) fit the equation $Z = 3 + 3X + 3Y$ with 10 data points

(5) fit the equation $Z = 4 + 4X + 4Y$ with 20 data points

(6) fit the equation $Z = 5 + 5X + 5Y$ with 30 data points

(7) fit the equation $Z = 3 - 3X - 3Y$ with 10 data points

(8) fit the equation $Z = 4 - 4X - 4Y$ with 20 data points

(9) fit the equation $Z = 5 - 5X - 5Y$ with 30 data points

From these results, we note that these two simple methods give similar results to the least squares method and these estimates from all three methods seem to be biased. The magnitude of the bias of estimates increases as σ_x , σ_y , and σ_z increase.

CHAPTER V

CONCLUSION

These two simple methods have been developed and extended to multiple regression with two independent variables. Proofs that the estimators are consistent are given. A small simulation study has been made to compare these methods to the least squares method.

From the simulated examples, it can be seen that the two simple methods give similar results to the least squares method. It appears that when the standard deviation of the error terms becomes larger, the absolute values of the estimates of b and c get smaller. Thus the estimates appear to be biased. It looks like that the magnitude of the bias of the estimates is a function of the standard deviation of the error terms. This might be the subject of a future study.

According to assumption (6), X and Y are uncorrelated, this is used to prove that the estimates are consistent. However, in practice, X and Y might be correlated. It would be interesting to try and prove that the estimates are still consistent when X and Y are correlated.

These two methods give an alternative to least squares in which the estimates are known to be consistent. However the numerical results obtained in simulation studies do not indicate that the methods are better than least squares.

BIBLIOGRAPHY

1. Graybill, F. A.: An Introduction to Linear statistical Models, McGraw-Hill, New York, 1961.

2. Wald, A.: " The Fitting of Straight lines if Both Variables Are Subject to Error," Ann. Math. Statist., vol. 11, pp. 284-300, 1940-1941.

3. Bartlett, M. S.: " Fitting a Straight Line when Both Variables Are Subject to Error," Biometrics, vol. 5, pp. 207-212, 1949.

4. Gibson, W. M. and G. H. Jowett: " Three-group Regression Analysis," Appl. Statist., vol. 6, pp. 114-122, 1957.

5. Draper, N. R. and H. Smith: Applied Regression Analysis, John Wiley & Sons Inc., New York, 1966.

6. Keeping, E. S.: " Note on Wald's Method of Fitting a Straight Line when Both Variables Are Subject to Error," Biometrics, vol. 12, pp. 445-448, 1956.

7. Berkson, J.: " Are There Two Regressions?" Am. Statist Assoc., vol. 45, pp. 164-180, 1950.

Appendix A

The proofs for the equations

in Section 3.6

The notations S_X^2, S_Y^2, S_Z^2, S_{XY}, S_{XZ}, and S_{YZ} for the mathematical variables and S_x^2, S_y^2, S_z^2, S_{xy}, S_{xz}, and S_{yz} for the observable variables are the same as those in Section 3.6. x, y, and z are the observable variables, while X, Y, and Z are the unobservable mathematical variables, $x = X + e_x$, $y = Y + e_y$, and $z = Z + e_z$ are random errors with zero means and σ_x^2, σ_y^2, and σ_z^2 variances. Note that X, Y, and Z are not considered as random variables in this appendix.

(1) The proof for (3-18-1), (3-18-2), and (3-18-3):

$$S_x^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

$$= \frac{\sum((X_i + e_{xi}) - (\bar{X} + \bar{e}_x))^2}{n - 1}$$

$$= \frac{\sum((X_i - \bar{X}) + (e_{xi} - \bar{e}_x))^2}{n - 1}$$

$$= \frac{\sum(X_i - \bar{X})^2}{n - 1} + \frac{\sum(e_{xi} - \bar{e}_x)^2}{n - 1} + 2\,\frac{\sum(X_i - \bar{X})(e_{xi} - \bar{e}_x)}{n - 1}$$

$$E(S_X^2) = E(S_X^2) + E\left(\frac{\sum(e_{xi} - \bar{e}_x)^2}{n - 1}\right) + 2E\left(\frac{\sum(X_i - \bar{X})(e_{xi} - \bar{e}_x)}{n - 1}\right)$$

Since X is mathematical variable,

$$E(S_X^2) = S_X^2$$

Since $E(e_x) = 0$ and $\text{var}(e_x) = \sigma_x^2$,

$$E\left(\frac{\sum(e_{xi} - \bar{e}_x)^2}{n - 1}\right) = \sigma_x^2$$

and
$$E\left(\frac{\sum(X_i - \bar{X})(e_{xi} - \bar{e}_x)}{n - 1}\right) = \frac{\sum(X_i - \bar{X})}{n - 1} E(e_{xi} - \bar{e}_x)$$

$$= \frac{\sum(X_i - \bar{X})}{n - 1} 0$$

$$= 0$$

Therefore

$$E(s_x^2) = s_X^2 + \sigma_x^2$$

Similarly

$$E(s_y^2) = s_Y^2 + \sigma_y^2$$

$$E(s_z^2) = s_Z^2 + \sigma_z^2$$

(2) The proof for (3-18-4), (3-18-5), and (3-18-6):

$$s_{xy} = \frac{\sum(x_i - \bar{x})(Y_i - \bar{y})}{n - 1}$$

$$= \frac{\sum((X_i + e_{xi}) - (\bar{X} + \bar{e}_x))((Y_i + e_{yi}) - (\bar{Y} + \bar{e}_y))}{n - 1}$$

$$= \frac{\sum((X_i - \bar{X}) + (e_{xi} - \bar{e}_x))((Y_i - \bar{Y}) + (e_{yi} - \bar{e}_y))}{n - 1}$$

$$= \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n - 1} + \frac{\sum(X_i - \bar{X})(e_{yi} - \bar{e}_y)}{n - 1}$$

$$+ \frac{\sum(Y_i - \bar{Y})(e_{xi} - \bar{e}_x)}{n - 1} + \frac{\sum(e_{xi} - \bar{e}_x)(e_{yi} - \bar{e}_y)}{n - 1}$$

$$E(S_{xy}) = E(S_{XY}) + E(\frac{\sum(X_i - \bar{X})(e_{yi} - \bar{e}_y)}{n - 1})$$

$$+ E(\frac{\sum(Y_i - \bar{Y})(e_{xi} - \bar{e}_x)}{n - 1}) + E(\frac{\sum(e_{xi} - \bar{e}_x)(e_{yi} - \bar{e}_y)}{n - 1})$$

Since X and Y are mathematical values, the equation becomes

$$E(S_{xy}) = S_{XY} + \frac{\sum(X_i - \bar{X})}{n - 1} E(e_{yi} - \bar{e}_y) + \frac{\sum(Y_i - \bar{Y})}{n - 1} E(e_{xi} - \bar{e}_x)$$

$$+ E(\frac{\sum(e_{xi} - \bar{e}_x)(e_{yi} - \bar{e}_y)}{n - 1})$$

Since $E(e_x) = 0$, $E(e_y) = 0$, and e_x and e_y are uncorrelated, the above equation becomes

$$E(S_{xy}) = S_{XY} + \frac{\sum(X_i - \bar{X})}{n - 1} 0 + \frac{\sum(Y_i - \bar{Y})}{n - 1} 0 + 0$$

$$E(S_{xy}) = S_{XY}$$

Similarly $E(S_{xz}) = S_{XZ}$, and $E(S_{yz}) = S_{YZ}$

(3) The proof for (3-18-7):

$$s_Z^2 = \frac{\sum(Z_i - \bar{Z})^2}{n - 1}$$

Substituting $(a + bX_i + cY_i)$ for Z_i, the equation becomes

$$s_Z^2 = \frac{\sum((a + bX_i + cY_i) - (a + b\bar{X} + c\bar{Y}))^2}{n - 1}$$

$$= \frac{\sum(b(X_i - \bar{X}) + c(Y_i - \bar{Y}))^2}{n - 1}$$

$$= \frac{\sum(b^2(X_i - \bar{X})^2 + 2bc(X_i - \bar{X})(Y_i - \bar{Y}) + c^2(Y_i - \bar{Y})^2)}{n - 1}$$

$$= \frac{b^2\sum(X_i - \bar{X})^2}{n - 1} + \frac{2bc\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n - 1} + \frac{c^2\sum(Y_i - \bar{Y})^2}{n - 1}$$

$$= b^2s_X^2 + 2bcs_{XY} + c^2s_Y^2$$

(4) The proof for (3-18-8) and (3-18-9):

$$s_{XZ} = \frac{\sum(X_i - \bar{X})(Z_i - \bar{Z})}{n - 1}$$

$$= \frac{\sum(X_i - \bar{X})((a + bX_i + cY_i) - (a + b\bar{X} + c\bar{Y}))}{n - 1}$$

$$= \frac{\sum(X_i - \bar{X})(b(X_i - \bar{X}) + c(Y_i - \bar{Y}))}{n - 1}$$

$$= \frac{b\sum(X_i - \bar{X})^2}{n - 1} + \frac{c\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

$$= bS_X^2 + cS_{XY}$$

Similarly

$$S_{YZ} = cS_Y^2 + bS_{XY}$$

Appendix B

FORTRAN IV program for fitting the equation $Z = a + bX + cY$

```
C       **************
C       MAIN PROGRAM
C       **************
C
C       **************
C       VARIABLE DESCRIPTION
C       NO AND NM                 ARE NUMBER OF DATA AND ITERATION
C       XX, YY, AND ZZ            ARE UNOBSERVABLE VARIABLES
C       X, Y, AND Z               ARE OBSERVABLE VARIABLES
C       EX, EY, AND EZ            ARE RANDOM ERRORS
C       STDX, STDY, AND STDZ      ARE STANDARD DEVIATIONS OF EX, EY, AND EZ
C       XAVE, YAVE, AND ZAVE      ARE AVERAGE VALUES OF X, Y, AND Z
C       **************
C
        DIMENSION X(70),Y(70),Z(70),XX(70),YY(70),ZZ(70),
       1A(10,10,3),B(10,10,3),C(10,10,3),AA(3),BB(3),CC(3)
        NSEED=75757575
C
        WRITE(6,500)
        DO 9 LL=1,3
C       **************
C       READ THE RANGE OF VARIABLES: XX AND YY
C       **************
        READ(5,103)A1,A2,B1,B2
C
C       **************
C       READ THE COEFFICIENTS OF THE EQUATION: ZZ = AA1 + BB2*XX + CC3*YY
C       **************
        READ(5,102)AA1,BB2,CC3
C
C       **************
C       READ THE MEAN, STD DEVIATION, NUMBER OF DATA AND ITERATION
C       **************
        READ(5,100)STDX,EXPX,STDY,EXPY,STDZ,EXPZ,NO,NM
        WRITE(6,400)A1,A2,B1,B2
        WRITE(6,300)AA1,BB2,CC3
        WRITE(6,401)NO
        DO 10 M=1,9
        DO 8 L=1,3
        A(M,NM,L)=0.
        B(M,NM,L)=0.
   8    C(M,NM,L)=0.
        DO 11 N=1,NM
C
C       **************
C       GENERATING OBSERVABLE VARIABLES: X,Y, AND Z
C       **************
        DO 12 I=1,NO
        XX(I)=(A1-A2)*RAN(NSEED)+A2
        EX=ATIME(EXPX,STDX,NSEED)
        X(I)=XX(I)+EX
        YY(I)=(B1-B2)*RAN(NSEED)+B2
        EY=ATIME(EXPY,STDY,NSEED)
        Y(I)=YY(I)+EY
```

```
      ZZ(I)=AA1+BB2*XX(I)+CC3*YY(I)
      EZ=STIMN(EXPZ,STDZ,NSEED)
   12 Z(I)=ZZ(I)+EZ
      XAVE=0.
      YAVE= .
      ZAVE=0.
      DO 13 I=1,NO
      XAVE=XAVE+X(I)
      YAVE=YAVE+Y(I)
   13 ZAVE=ZAVE+Z(I)
      XAVE=XAVE/FLOAT(NO)
      YAVE=YAVE/FLOAT(NO)
      ZAVE=ZAVE/FLOAT(NO)
      CALL REARR(X,Y,Z,XX,YY,ZZ,NO)
      DO 14 L=1,3
      GO TO (51,51,50),L
C
C     **************
C     CALL LEAST SQUARE METHOD
C     **************
   50 CALL LSM(X,Y,Z,XAVE,YAVE,ZAVE,NO,C00,C11,C22)
      AA(L)=C00
      BB(L)=C11
      CC(L)=C22
      GO TO 53
   51 IF(L .EQ. 1) GO TO 40
      GO TO 41
C
C     **************
C     CALL TWO GROUP METHOD
C     **************
   40 CALL TAT(X,Y,Z,XAVE,YAVE,ZAVE,NO,C00,C11,C22,L)
      AA(L)=C00
      BB(L)=C11
      CC(L)=C22
      GO TO 53
C
C     **************
C     CALL THREE GROUP METHOD
C     **************
   41 CALL TAT(X,Y,Z,XAVE,YAVE,ZAVE,NO,C00,C11,C22,L)
      AA(L)=C00
      BB(L)=C11
      CC(L)=C22
   53 A(M,NM,L)=A(M,NM,L)+AA(L)
      B(M,NM,L)=B(M,NM,L)+BB(L)
   14 C(M,NM,L)=C(M,NM,L)+CC(L)
   11 CONTINUE
      DO 54 L=1,3
      A(M,NM,L)=A(M,NM,L)/FLOAT(NM)
      B(M,NM,L)=B(M,NM,L)/FLOAT(NM)
   54 C(M,NM,L)=C(M,NM,L)/FLOAT(NM)
      WRITE(6,402)STDX,(A(M,NM,L),B(M,NM,L),C(M,NM,L),L=1,3)
      STDX=STDX+.1
      STDY=STDY+.1
   10 STDZ=STDZ+.1
      WRITE(6,403)
    9 CONTINUE
  100 FORMAT(6F5.1,2I3)
  102 FORMAT(3F5.2)
  103 FORMAT(4F5.2)
  300 FORMAT( 9X,'A =',F4.1,7X,'B =',F4.1,7X,'C =',F4.1)
  400 FORMAT(6X,'REGION OF X IS',F6.2,2X,'TO',F6.2,4X,'REGION OF Y IS',F
     16.2,2X,'TO',F6.2)
  401 FORMAT(/6X,'NO =',I3,4X,'TWO GROUP',12X,'THREE GROUP',11X,'LEAST S
     1QUARE'//6X,'STD',3(5X,'A',6X,'B',6X,'C',2X)/)
  402 FORMAT(6X,F3.1,2X,3(F6.2,1X,F6.2,1X,F6.2,2X))
  403 FORMAT(//)
  600 FORMAT(1H1)
      STOP
      END
      SUBROUTINE TAT(X,Y,Z,XAVE,YAVE,ZAVE,NO,A,B,C,L)
C     ***************
C     TWO GROUR METHOP AND THREE GROUP METHOD FOR THREE DIMENSIONS
C     ***************
C
C     ***************
C     L = 1 IS FOR TWO GROUP METHOD
C     L = 2 IS FOR THREE GROUP METHOD
C     NO    = THE NUMBER OF OBSERVATION
C     XAVE = THE AVERAGE OF X OBSERVATION
C     YAVE = THE AVERAGE OF Y OBSERVATION
C     ZAVE = THE AVERAGE OF Z OBSERVATION
C     RE1 = THE RESIDUAL VALUES OF Y ON X
C     RE2 = THE RESIDUAL VALUES OF Z ON X
C     ***************
C
      DIMENSION X(70),Y(70),Z(70),RE1(70),RE2(70),DA(70)
      DATA DA/70*0./
      GO TO (40,41),L
   40 NX1=2
      NX2=2
      GO TO 42
   41 NX1=3
      NX2=3
   42 LU1=NO/NX1
C
C     **************
C     FIND RELATIONSHIP BETWEEN X AND Z
C     **************
      XAV1=0.
      ZAV1=0.
      DO 15 I=1,LU1
      XAV1=XAV1+X(I)
   15 ZAV1=ZAV1+Z(I)
      XAV1=XAV1/FLOAT(LU1)
      ZAV1=ZAV1/FLOAT(LU1)
      XAV3=0.
      ZAV3=0.
      LU2=NO/NX2
      MMX=NO-LU2+1
      DO 16 I=MMX,NO
      XAV3=XAV3+X(I)
   16 ZAV3=ZAV3+Z(I)
      XAV3=XAV3/FLOAT(LU2)
      ZAV3=ZAV3/FLOAT(LU2)
      FXB=(ZAV3-ZAV1)/(XAV3-XAV1)
      FXA=ZAVE-FXB*XAVE
C
C     **************
C     FIND RELATIONSHIP BETWEEN X AND Y
C     **************
      YAV1=0.
      DO 17 I=1,LU1
   17 YAV1=YAV1+Y(I)
```

```
      YAV1=YAV1/FLOAT(LU1)
      YAV3=0.
      DO 18 I=MMX,NO
   18 YAV3=YAV3+Y(I)
      YAV3=YAV3/FLOAT(LU2)
      SXB=(YAV3-YAV1)/(XAV3-XAV1)
      SXA=YAV3-SXB*XAV3
C
C     **************
C     FIND RELATIONSHIP BETWEEN TWO RESIDUALS
C     **************
      DO 20 I=1,NO
      RE2(I)=Z(I)-FXA-FXB*X(I)
   20 RE1(I)=Y(I)-SXA-SXB*X(I)
      CALL REARR(RE1,RE2,DA,DA,DA,DA,NO)
      R1V1=0.
      R1V3=0.
      R2V1=0.
      R2V3=0.
      DO 30 I=1,LU1
      R1V1=R1V1+RE1(I)
   30 R2V1=R2V1+RE2(I)
      R1V1=R1V1/FLOAT(LU1)
      R2V1=R2V1/FLOAT(LU1)
      MMX=NO-LU1+1
      DO 31 I=MMX,NO
      R1V3=R1V3+RE1(I)
   31 R2V3=R2V3+RE2(I)
      R2V3=R2V3/FLOAT(LU2)
      R1V3=R1V3/FLOAT(LU2)
      TXB=(R2V3-R2V1)/(R1V3-R1V1)
      CC0=FXA-TXB*SXA
      CC1=FXB-TXB*SXB
      CC2=TXB
C
C     **************
C     OBTAIN THE ESTIMATORS FOR THE COEFFICIENTS OF EQUATION: A, B, AND C
C     **************
      A=FXA-TXB*SXA
      B=FXB-TXB*SXB
      C=TXB
      RETURN
      END
      SUBROUTINE REARR(F1,F2,F3,F4,F5,F6,NO)
C     **************
C     REARRAGE F1 VARIABLE, LET ITS VALUES INCREASE IN ORDER.
C     F2,F3,F4,F5,AND F6 CORRESPOND TO F1, THESE CORRESPONDING VARIABLES
C     WILL BE REARRAGED ACCORDING TO THE NEW ORDER OF F1
C     **************
C
      DIMENSION F1(70),F2(70),F3(70),F4(70),F5(70),F6(70)
      K=1
      F1(NO+1)=10000.
      DO 20 M=1,NO
      DO 21 I=K,NO
      IF(F1(M)-F1(I+1))21,21,22
   22 STO1=F1(I+1)
      STO2=F2(I+1)
      STO3=F3(I+1)
      STO4=F4(I+1)
      STO5=F5(I+1)
      STO6=F6(I+1)
      F1(I+1)=F1(M)
      F2(I+1)=F2(M)
      F3(I+1)=F3(M)
      F4(I+1)=F4(M)
      F5(I+1)=F5(M)
      F6(I+1)=F6(M)
      F1(M) = STO1
      F2(M)=STO2
      F3(M)=STO3
      F4(M)=STO4
      F5(M)=STO5
      F6(M)=STO6
   21 CONTINUE
   20 K=K+1
      RETURN
      END
```

```
LEVEL  19                    LSM                    DATE = 72342              16/12/49

      SUBROUTINE LSM(X,Y,Z,XAVE,YAVE,ZAVE,NO,A,B,C)
C     **************
C     LEAST SQUARE METHOD FOR THREE DIMENSIONS; Z = A + BX + CY
C     **************
C
      DIMENSION X(70),Y(70),Z(70)
      SX=0.
      SY=0.
      SXY=0.
      SXZ=0.
      SYZ=0.
      DO 22 I=1,NO
      SX=(X(I)-XAVE)**2+SX
      SY=(Y(I)-YAVE)**2+SY
      SXY=(X(I)-XAVE)*(Y(I)-YAVE)+SXY
      SXZ=(X(I)-XAVE)*(Z(I)-ZAVE)+SXZ
   22 SYZ=(Y(I)-YAVE)*(Z(I)-ZAVE)+SYZ
      SS=SX*SY-(SXY)**2
      C   =SX/SS*SYZ-SXY/SS*SXZ
      B   =SY/SS*SXZ-SXY/SS*SYZ
      A   = ZAVE-B*XAVE-C*YAVE
      RETURN
      END
```

```
LEVEL  19                    ATIME                  DATE = 72342              16/12/49

      FUNCTION ATIME(P1,P2,NSEED)
C     *************
C     NORMAL DIST. RANDOM NUMBER GENERATION
C     *************
C
C     *************
C     P1 = MEAN OF NORMAL DIST.
C     P2 = STD. DEVIATION OF NORMAL DIST.
C     *************
C
      SUM=0.
      DO 101 I=1,12
  101 SUM=SUM+RAN(NSEED)
      ATIME=(SUM-6.)*P2+P1
      RETURN
      END
```

```
LEVEL  19                    RAN                    DATE = 72342              16/12/49

      FUNCTION RAN(NSEED)
C     **************
C     RANDOM NUMBER GENERATION
C     **************
C
      NSEED=IABS(NSEED*655393)
      RAN=FLOAT(MOD(NSEED,33554432))/FLOAT(33554432)
      RETURN
```

Appendix C

The results of simulation study in Chapter IV

```
    REGION OF X IS 15.00   TO   2.00     REGION OF Y IS 20.00   TO   2.00
    A = 3.0           B = 4.0          C = 2.0

NO = 10      TWO GROUP                 THREE GROUP               LEAST SQUARE

STD       A        B        C         A        B        C         A        B        C

0.1      1.84     4.06     2.03      2.35     4.02     2.01      2.30     4.03     2.01
0.2      2.32     4.11     1.95      2.72     4.05     1.98      2.51     4.09     1.96
0.3      5.77     3.58     2.01      4.92     3.71     2.01      5.16     3.70     1.99
0.4      3.57     3.89     2.00      2.44     3.99     1.93      3.30     3.92   · 2.00
0.5      2.30     4.05     2.07      3.26     4.00     1.99      3.28     3.95     2.04
0.6      3.23     3.86     2.14      3.54     3.94     2.03      4.07     3.90     2.02
0.7      5.55     3.79     1.97      3.17     3.95     2.10      3.85     3.99     1.98
0.8      4.99     3.41     2.03      4.82     3.51     2.15      6.06     3.50     2.05
0.9      5.03     4.01     1.57      6.68     3.93     1.67      7.43     3.66     1.63


    REGION OF X IS 22.00   TO   5.00     REGION OF Y IS 14.00   TO   1.00
    A =-3.0           B = 3.0          C =-4.0

NO = 20      TWO GROUP                 THREE GROUP               LEAST SQUARE

STD       A        B        C         A        B        C         A        B        C

0.1     -3.31     3.01    -3.97     -3.17     3.01    -4.00     -3.14     3.01    -4.00
0.2     -2.94     3.01    -4.02     -2.51     2.97    -4.02     -2.65     2.98    -4.01
0.3     -4.84     3.04    -3.83     -4.40     3.04    -3.90     -4.45     3.05    -3.92
0.4     -2.27     2.93    -3.96     -3.18     2.97.   -3.91     -3.25     2.99    -3.95
0.5     -2.80     2.86    -3.76     -1.89     2.82    -3.80     -2.01     2.83    -3.79
0.6     -0.34     2.80    -4.01     -0.60     2.79    -3.95     -0.80     2.81    -3.96
0.7     -1.16     2.77    -3.84     -3.64     2.90    -3.77     -2.66     2.82    -3.75
0.8     -4.37     2.88    -3.60     -6.63     2.94    -3.43     -4.82     2.90    -3.59
0.9     -6.01     3.13    -3.86     -7.00     3.20    -3.87 .   -7.38     3.15    -3.72


    REGION OF X IS 13.00   TO   2.00     REGION OF Y IS 18.00   TO   4.00
    A = 5.0           B =-4.0          C = 2.0

NO = 30      TWO GROUP                 THREE GROUP               LEAST SQUARE

STD       A        B        C         A        B        C         A        B        C

0.1      4.59    -4.00     2.04      4.31    -3.99     2.04      4.55    -3.99     2.03
0.2      5.12    -4.02     1.93      5.10    -4.01     1.99      4.98    -3.98     1.99
0.3      5.25    -4.08     2.03      5.36    -4.04     1.98      5.23    -4.02     2.00
0.4      4.14    -3.87     1.99      3.84    -3.89     2.03      4.08    -3.88     2.00
0.5      5.38    -4.00     2.02      5.16    -3.96     1.99      5.20    -3.94     1.97
0.6      5.11    -3.81     1.84      5.11    -3.84     1.86      5.14    -3.78     1.82
0.7      4.36    -3.87     2.00      3.83    -3.89     2.05      4.09    -3.84     2.01
0.8      5.80    -4.09     1.96      5.11    -3.94     1.93      3.77    -3.78     1.94
0.9      1.20    -3.43     1.36      1.69    -3.55     2.00      1.62    -3.51     1.97
```

```
   REGION OF X IS 25.00  TO 10.00    REGION OF Y IS 17.00  TO  5.00
   A = 3.0         B = 3.0         C = 3.0

NO = 10     TWO GROUP               THREE GROUP              LEAST SQUARE

STD       A        B        C        A        B        C        A        B        C

0.1     1.57     3.04     3.14     2.06     3.02     3.03     2.23     3.02     3.02
0.2     2.26     3.07     2.93     3.57     3.02     2.90     2.65     3.05     2.93
0.3     9.33     2.67     3.07     6.47     2.74     3.05     7.07     2.75     2.98
0.4     5.13     2.99     2.96     4.34     2.96     2.87     4.46     2.92     2.97
0.5     2.22     3.02     3.07     2.52     3.00     3.07     3.47     2.99     3.00
0.6     5.56     2.94     3.04     3.65     2.97     3.00     5.07     2.92     2.96
0.7     8.36     2.78     2.18     6.08     2.88     2.95     6.32     2.93     2.82
0.8    11.10     2.49     3.02     6.82     2.54     3.14    10.24     2.59     2.93
0.9     9.27     3.11     2.57     8.72     2.90     2.58     9.19     2.93     2.47
```

```
   REGION OF X IS 20.00  TO  4.00    REGION OF Y IS 15.00  TO  3.00
   A = 4.0         B = 4.0         C = 4.0

NO = 20     TWO GROUP               THREE GROUP              LEAST SQUARE

STD       A        B        C        A        B        C        A        B        C

0.1     4.07     3.98     4.02     3.88     3.98     4.03     3.85     3.99     4.02
0.2     3.38     4.03     4.04     3.63     4.00     4.05     3.93     3.99     4.02
0.3     5.86     3.95     3.91     4.58     4.03     3.95     4.43     3.97     4.04
0.4     4.55     4.00     3.96     5.99     3.94     3.89     5.99     3.92     3.92
0.5     6.55     3.90     3.87     8.21     3.87     3.74     7.41     3.89     3.79
0.6     5.71     3.92     3.94     5.14     3.97     3.94     6.26     3.89     3.91
0.7     7.22     4.18     3.37     5.55     4.11     3.63     6.04     4.11     3.57
0.8    11.91     3.62     3.65    12.57     3.61     3.60    12.73     3.59     3.61
0.9     8.35     3.85     3.78     9.76     3.81     3.67    11.37     3.77     3.55
```

```
   REGION OF X IS 12.00  TO  1.00    REGION OF Y IS 12.00  TO  2.00
   A = 5.0         B = 5.0         C = 5.0

NO = 30     TWO GROUP               THREE GROUP              LEAST SQUARE

STD       A        B        C        A        B        C        A        B        C

0.1     4.95     5.04     4.95     5.11     5.03     4.95     5.19     5.02     4.94
0.2     4.85     5.02     5.00     4.89     4.98     5.03     5.14     4.97     5.00
0.3     6.07     4.95     4.89     5.23     4.97     4.98     5.73     4.93     4.95
0.4     6.23     5.24     4.54     6.01     5.23     4.68     6.04     5.15     4.76
0.5     8.95     4.84     4.64     8.60     4.85     4.70     9.09     4.86     4.76
0.6     7.49     4.73     4.97     6.79     4.74     4.96     7.27     4.68     4.94
0.7     9.18     4.78     4.62     8.51     4.85     4.66     8.91     4.85     4.61
0.8     8.80     4.96     4.56    10.97     4.73     4.45    11.37     4.67     4.45
0.9     6.05     5.07     4.73     7.78     4.98     4.57     8.92     4.84     4.53
```

```
 REGION OF X IS 20.00  TO  3.00    REGION OF Y IS 20.00  TC  3.00
 A = 3.0          B =-3.0          C =-3.0


NO = 10    TWO GROUP              THREE GROUP           LEAST SQUARE

STD       A        B        C        A        B        C        A        B        C

0.1     3.38   -3.01   -3.00    3.33   -3.00   -3.01    3.33   -3.00   -3.00
1.1    -6.35   -2.71   -2.45    0.73   -3.21   -2.41   -0.46   -3.05   -2.52
2.1    -6.25   -1.79   -3.27  -10.73   -1.69   -2.87   -9.40   -1.65   -3.06
3.1   -12.41   -2.61   -2.10   -9.39   -2.54   -2.44  -11.84   -2.48   -2.29
4.1   -15.55   -2.56   -1.63  -20.79   -1.92   -2.03  -20.33   -1.98   -2.00
5.1   -38.22   -1.68   -0.94  -24.03   -2.22   -1.61  -32.01   -1.54   -1.51
6.1   -32.93   -1.79   -1.51  -30.30   -1.16   -1.62  -33.00   -1.11   -1.42
7.1   -51.60   -0.00   -1.18  -55.44   -0.57   -0.89  -49.20   -0.45   -0.97
8.1   -38.20   -0.86   -1.36  -27.66   -1.22   -1.77  -38.18   -1.21   -1.03


  REGION OF X IS 20.00  TO  3.00    REGION OF Y IS 20.00  TC  3.00
  A = 4.0          B =-4.0          C =-4.0


NO = 20    TWO GROUP              THREE GROUP           LEAST SQUARE

STD       A        B        C        A        B        C        A        B        C

0.1     4.08   -3.98   -4.03    4.13   -3.98   -4.03    4.16   -3.99   -4.02
1.1     1.45   -3.93   -3.83    1.66   -3.94   -3.85    2.23   -3.90   -3.93
2.1   -12.81   -3.69   -3.04  -16.76   -3.27   -3.09  -13.35   -3.32   -3.35
3.1   -22.71   -3.02   -2.76  -27.97   -2.67   -2.66  -24.69   -2.96   -2.67
4.1   -38.38   -2.39   -2.08  -35.23   -2.54   -2.12  -38.99   -2.27   -2.09
5.1   -45.83   -2.12   -1.80  -44.03   -2.09   -1.93  -48.19   -1.94   -1.71
6.1   -69.41   -1.82   -0.11  -62.24   -1.81   -0.63  -62.76   -1.79   -0.61
7.1   -61.97   -1.57   -0.82  -64.89   -1.44   -0.72  -64.40   -1.54   -0.70
8.1   -59.33   -1.09   -1.44  -64.65   -0.73   -1.34  -63.14   -0.74   -1.51


  REGION OF X IS 20.00  TO  3.00    REGION OF Y IS 20.00  TC  3.00
  A = 5.0          B =-5.0          C =-5.0


NO = 30    TWO GROUP              THREE GROUP           LEAST SQUARE

STD       A        B        C        A        B        C        A        B        C

0.1     4.96   -5.02   -4.97    4.86   -5.01   -4.97    4.76   -5.01   -4.96
1.1     4.31   -4.95   -5.03    3.43   -4.90   -4.92    1.78   -4.81   -4.88
2.1    -8.09   -4.29   -4.49   -4.88   -4.52   -4.55   -8.38   -4.26   -4.49
3.1   -20.31   -4.69   -3.05  -24.17   -4.32   -3.03  -28.60   -4.06   -2.96
4.1   -41.73   -3.22   -2.74  -46.95   -3.18   -2.30  -47.92   -2.82   -2.60
5.1   -43.57   -2.74   -3.05  -44.37   -2.83   -2.91  -44.67   -2.76   -2.95
6.1   -53.76   -2.37   -2.04  -52.14   -2.67   -2.02  -58.97   -2.21   -1.90
7.1   -57.39   -2.03   -2.39  -63.51   -1.87   -2.07  -64.52   -1.91   -1.94
8.1   -71.55   -1.39   -1.39  -78.86   -1.73   -1.11  -77.09   -1.77   -1.19
```