E82-10270
NASA-CR-167597

AgRISTARS

IT-L2-04228
JSC-17814

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Inventory Technology Development

March 1982

GENERAL MULTIYEAR AGGREGATION TECHNOLOGY: METHODOLOGY AND SOFTWARE DOCUMENTATION

T. C. Baker

(E82-10270) GENERAL MULTIYEAR AGGREGATION TECHNOLOGY: METHODOLOGY AND SOFTWARE DOCUMENTATION (Lockheed Engineering and Management) 45 p HC A03/MF A01 CSCL 02C N82-24548
Unclas
G3/43 00270

Lockheed Engineering and Management Services Company, Inc.
1830 NASA Road 1, Houston, Texas 77258

Lyndon B. Johnson Space Center
Houston, Texas 77058

IT-L2-04228
JSC-17814

GENERAL MULTIYEAR AGGREGATION TECHNOLOGY:

METHODOLOGY AND SOFTWARE DOCUMENTATION

Job Order 72-422

This report describes the technology development activities of the Inventory Technology Development project of the AgRISTARS program.

PREPARED BY

T. C. Baker

APPROVED BY

D. E. Phinney, Project Manager
Technology Development Office

M. D. Pore, Supervisor
Statistical Analysis Section

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.

Under Contract NAS 9-15800

For

Earth Resources Applications Division
Space and Life Sciences Directorate

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

March 1982

LEMSCO-17153

PRECEDING PAGE BLANK NOT FILMED

CONTENTS

TABLES

FIGURES

1. INTRODUCTION

This paper presents a general methodology for estimating a stratum's (observable) crop acreage proportion for a specified season of a target year from the crop's estimated acreage proportion for sample segments from within the stratum. Sample segment data for several years and seasons are generally used in conjunction with those for the target year and season. The proposed methodology is an application of estimation from a mixed Analysis of Variance (AOV) model.

The specific model proposed is described and developed in section 2 of this document. A general discussion of estimation using the model is also presented in this section. In section 3, the use of the model in crop acreage proportion estimation is described. The general applicability of the development is illustrated by three examples. A documentation of the Statistical Analysis System (SAS) implementation of the methodology and sample runs for the examples of section 3 are presented in section 4.

In the following discussion, matrix notation is used extensively. In particular, capital letters (with or without subscripts) refer to matrices. Underscored small and greek letters (with or without subscripts) refer to column vectors. Small and greek letters that are used with subscripts and without underscores refer to elements of vectors and matrices. Constants are denoted as small or greek letters with neither subscripts nor underscores. The transpose of a matrix, say A, is denoted A'. The n-by-n identity matrix is indicated by I_n. If A is a square matrix, A^- denotes any matrix such that $AA^-A = A$. For A, n-by-n nonsingular, A^{-1} denotes a matrix such that $AA^{-1} = A^{-1}A = I_n$. The symbols $\underline{0}_n$ and $\underline{1}_n$ represent column vectors of length n composed entirely of zero's and one's, respectively. When the dimensions of a matrix or vector are not specifically noted, it is assumed that they are conformable with the designated operations. Mathematical expectation is denoted E[ ] for both matrix and scalar operands. Finally, if f( ) is a function and $\underline{p}$ is a vector the elements of which are denoted p_i, then $\underline{f}(\underline{p})$ denotes a vector such that $f_i = f(p_i)$.

2. THE BASIC MODEL AND ITS USE IN ESTIMATION

The proposed methodology is an application of the mixed AOV model:

$$\underline{y} = X\underline{\beta} + Z\underline{r} + \underline{e} \tag{2.1}$$

where

$\underline{y}$ = an n-by-1 response vector

$\underline{\beta}$ = a t-by-1 vector of fixed-effects coefficients

X = an n-by-t matrix representing the fixed-effects design

$\underline{r}$ = an s-by-1 vector of random-effect coefficients

Z = an n-by-s matrix representing the random-effect design

$\underline{e}$ = an n-by-1 vector of unexplained errors

The usual assumptions regarding this model are as follows:

a. $\underline{r}$ is a random vector with mean $\underline{0}_s$ and variance covariance (V-C) matrix V_r.

b. $\underline{e}$ is a random vector with mean $\underline{0}_n$ and V-C matrix V_e.

c. $\underline{r}$ and $\underline{e}$ are independent.

The particular application considered in this document has the property that $\underline{y}$ is a transformation of an observable vector $\underline{p}$; that is, $\underline{y} = \underline{f}(\underline{p})$ for some function f(•). Hence, in the following paragraphs, $\underline{p}$ is referred to as the response vector and $\underline{y}$ is referred to as the transformed-response vector. The elements of the response vector $\underline{p}$ are assumed to have the following properties for each i(i = 1, 2, •••, n):

d. $0 \leq p_i \leq 1$.

e. $E[p_i] = \pi_i$.

f. $Var[p_i] = E\left[(p_i - \pi_i)^2\right] \propto \pi_i(1 - \pi_i)$.

The following additional properties are assumed also.

g. V_e is a diagonal matrix with the i^{th} diagonal entry proportional to σ_ε^2 Var[f(p_i)] for some $\sigma_\varepsilon^2 > 0$.

h. V_r is a diagonal matrix with the i^{th} diagonal entry equal to $\gamma\sigma_\varepsilon^2$ for some $\gamma > 0$.

i. X and Z are both of full column rank [although (X:Z) need not be of full column rank].

j. n > rank (X:Z) + 1.

The choice of a transformation function f(•) will also be restricted to one of the three forms:

k. 'identity', $f(p_i) = p_i$.

l. 'log', $f(p_i) = \ln(p_i)$.

m. 'logit', $f(p_i) = .5 \times \ln[p_i/(1 - p_i)]$.

2.1 DEVELOPMENT OF THE ESTIMATION EQUATIONS

Under the conditions imposed above, the mean vector and V-C matrix, respectively, for $\underline{y}$ are

$$E[\underline{y}] = X\underline{\beta} = \underline{\mu} \tag{2.2}$$

and

$$Var[\underline{y}] = E[(\underline{y} - \underline{\mu})(\underline{y} - \underline{\mu})'] = V_e + ZV_rZ' \tag{2.3}$$

By denoting $V_e = W^{-1}\sigma_\varepsilon^2$ and $V_r = I_s\sigma_\varepsilon^2\gamma$, equation (2.3) becomes

$$Var[\underline{y}] = (W^{-1} + \gamma ZZ')\sigma_\varepsilon^2 = V^{-1}\sigma_\varepsilon^2 \tag{2.4}$$

Hence, the least squares estimator of $\underline{\beta}$ (given W and γ) is

$$\hat{\underline{\beta}} = (X'VX)^{-1}XV\underline{y} \tag{2.5}$$

and its V-C matrix is

$$\mathrm{Var}\left[\hat{\underline{\beta}}\right] = (X'VX)^{-1}\sigma_\varepsilon^2 \tag{2.6}$$

If C is a q-by-t matrix of estimable contrasts, then an unbiased estimate of $\underline{\gamma}_c = C\underline{\beta}$ is

$$\hat{\underline{\gamma}}_c = C\hat{\underline{\beta}} \tag{2.7}$$

and

$$\mathrm{Var}\left[\hat{\underline{\gamma}}_c\right] = C(X'VX)^{-1}C'\sigma_\varepsilon^2 = V_c \tag{2.8}$$

Since the quantities W, γ, and σ_ε^2 are usually unknown, equations (2.7) and (2.8) can only be used for estimation from the mixed model (2.1) after substitution of their respective estimates, $\hat{W}$, $\hat{\gamma}$, and $\hat{\sigma}_\varepsilon^2$. (For a discussion of the effect of this substitution on the asymptotic properties of the least squares estimator of $\underline{\gamma}_c$, shown in equation (2.7), see reference 1.)

2.1.1 ESTIMATION OF W, σ_ε^2, AND γ

Note that $W = V_e^{-1}\sigma_\varepsilon^2$ where the i^{th} diagonal entry of V_e is

$$v_{e_{ii}} = \mathrm{Var}[f(p_i)] = \sigma_f^2(\pi_i)\sigma_\varepsilon^2 \tag{2.9}$$

hence, $W = [w_{ij}]$ where

$$w_{ij} = \begin{cases} 1/\sigma_f^2(\pi_i) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{2.10}$$

Since $E[\underline{p}] = \underline{\pi}$, an initial estimate of W can be obtained by replacing each π_i in equation (2.10) by its respective estimate p_i.

That is, take $\hat{W} = \left[\hat{w}_{ij}\right]$ where

$$\hat{w}_{ij} = \begin{cases} 1/\sigma_f^2(p_i) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{2.11}$$

If equation (2.1) is premultiplied by $\hat{W}^{1/2}$ where $\hat{W} = (\hat{W}^{1/2})^2$, the result is

$$\hat{W}^{1/2}\underline{y} = \hat{W}^{1/2}X\underline{\beta} + \hat{W}^{1/2}Z\underline{r} + \hat{W}^{1/2}\underline{e} = \hat{W}^{1/2}(X:Z)\begin{pmatrix}\underline{\beta} \\ \cdots \\ \underline{r}\end{pmatrix} + \underline{e}_W \tag{2.12}$$

All the assumptions of model (2.1) hold for equation (2.12) so that now

$$\text{Var}[\underline{e}_W] = \hat{W}^{1/2}W^{-1}\hat{W}^{1/2}\sigma_\varepsilon^2 = I_n\sigma_\varepsilon^2 \tag{2.13}$$

Therefore, an initial estimate of σ_ε^2 is the residual mean square after fitting $\underline{\beta}$ and $\underline{r}$ in equation (2.12)

$$\hat{\sigma}_\varepsilon^2 = \underline{y}'\hat{W}\left|I_n - (X:Z)[(X:Z)'\hat{W}(X:Z)]^-(X:Z)'\hat{W}\right|\underline{y}/[n - \text{rank}(X:Z)] \tag{2.14}$$

which is unbiased for σ_ε^2 when $\hat{W} = W$.

The estimates of W and σ_ε^2 can be refined iteratively as follows. First, use the estimates of $\underline{\beta}$ and $\underline{r}$ from equation (2.12) to determine an estimated transformed-response vector $\hat{\underline{y}}$ for the model, that is

$$\hat{\underline{y}} = (X:Z)[(X:Z)'\hat{W}(X:Z)]^-(X:Z)'\hat{W}\underline{y} \tag{2.15}$$

Then define $\hat{\underline{p}} = \left[\hat{p}_i\right]$ where

$$\hat{p}_i = f^{-1}\left(\hat{y}_i\right) \tag{2.16}$$

After replacing each p_i in equation (2.11) with its corresponding $\hat{p}_i$ from equation (2.16), the entire procedure can be repeated to produce refined estimates of W and σ_ε^2. The iteration process can be continued until $\hat{W}$ (and hence $\hat{\sigma}_\varepsilon^2$) "stabilizes." In the following paragraphs, $\hat{W}$ and $\hat{\sigma}_\varepsilon^2$ denote the final stable estimates of W and σ_ε^2, respectively.

After $\hat{W}$ and $\hat{\sigma}_\varepsilon^2$ are available, γ can be estimated via the variance component analysis procedure known as Henderson's Method 3 (ref. 2). The resulting estimator for γ is

$$\hat{\gamma} = \frac{\left(\underline{y}'\hat{W}\{(X:Z)[(X:Z)'\hat{W}(X:Z)]^-(X:Z)' - X(X'\hat{W}X)^{-1}X'\}\hat{W}\underline{y}\Big/\hat{\sigma}_\varepsilon^2\right) - \text{rank }(X:Z) + s}{\text{trace}\left\{Z'\hat{W}\left[I_n - X(X'\hat{W}X)^{-1}X'\hat{W}\right]Z\right\}} \tag{2.17}$$

which is unbiased for γ when $\hat{W} = W$.

The estimator $\hat{\sigma}_\varepsilon^2$ in equation (2.14) has the property that $\hat{\sigma}_\varepsilon^2 > 0$ as required. However, the estimator $\hat{\gamma}$ in equation (2.17) may be either positive or negative. In the SAS implementation of this methodology, any $\hat{\gamma} < 0$ results in the termination of processing, since theoretically $\gamma > 0$.

2.1.2 INVERSE ESTIMATION

Recall that $\underline{y}$ represents a transformation of an observable response vector, $\underline{p}$. Hence the primary interest in estimation may not be

$$\underline{y}_c = C\underline{\beta} \tag{2.18}$$

but

$$\underline{a} = \underline{f}^{-1}(\underline{y}_c) \tag{2.19}$$

A straightforward estimator of $\underline{a}$ which employs equation (2.7) is

$$\hat{\underline{a}} = \underline{f}^{-1}\left(\hat{\underline{y}}_c\right) \tag{2.20}$$

Since $f(\cdot)$ may not be linear, the estimator $\hat{\underline{a}}$ in equation (2.20) may be biased for $\underline{a}$ and its exact matrix of mean squares (MMS) may be very difficult to determine. However, reasonable estimates of the bias and MMS for $\hat{\underline{a}}$ can be obtained as follows. [This development parallels that shown by Sielken and Dahn (ref. 3)].

Consider the approximation of each $\hat{a}_i = f^{-1}\left(\hat{y}_{c_i}\right)$ as a third order Taylor series expanded about y_{c_i} so that

$$\hat{y}_{c_i} = f^{-1}\left(y_{c_i}\right) + \left(\hat{y}_{c_i} - y_{c_i}\right) \frac{df^{-1}\left(\hat{y}_{c_i}\right)}{d\hat{y}_{c_i}} + 1/2 \left(\hat{y}_{c_i} - y_{c_i}\right)^2 \frac{d^2 f^{-1}\left(\hat{y}_{c_i}\right)}{d\hat{y}^2_{c_i}} \tag{2.21}$$

If $E\left[\hat{y}_{c_i}\right] = y_{c_i}$, taking the expectation of both sides of equation (2.21) yields

$$E\left[\hat{a}_i\right] = a_i + (1/2) \operatorname{Var}\left[\hat{y}_{c_i}\right] \frac{d^2 f^{-1}\left(\hat{y}_{c_i}\right)}{d\hat{y}^2_{c_i}} = a_i + .5 \times v_{c_{ii}} f^{(-2)}\left(\hat{y}_{c_i}\right) \tag{2.22}$$

so that

$$\operatorname{bias}\left[\hat{a}_i\right] = E\left[\hat{a}_i\right] - a_i = .5 \times v_{c_{ii}} f^{(-2)}\left(\hat{y}_{c_i}\right) \tag{2.23}$$

The matrix representation is

$$\operatorname{bias}\left[\underline{\hat{a}}\right] = \begin{bmatrix} v_{c_{11}} & 0 & \cdots & 0 \\ 0 & v_{c_{22}} & & \vdots \\ \vdots & & \ddots & \\ 0 & & \cdots & v_{c_{qq}} \end{bmatrix} \underline{f}^{(-2)}\left(\hat{y}_c\right) \tag{2.24}$$

To determine an estimate of MMS for $\underline{\hat{a}}$, drop the third term from equation (2.21) so that

$$\hat{a}_i = f^{-1}\left(y_{c_i}\right) + \left(\hat{y}_{c_i} - y_{c_i}\right) \frac{df^{-1}\left(\hat{y}_{c_i}\right)}{d\hat{y}_{c_i}} = a_i + \left(\hat{y}_{c_i} - y_{c_i}\right) f^{(-1)}\left(\hat{y}_{c_i}\right) \tag{2.25}$$

Then if $E\left[\hat{y}_{c_i}\right] = y_{c_i}$

$$\text{MMS}\left[\underline{\hat{\pi}}\right] = E\left[\left(\underline{\hat{\pi}} - \underline{\pi}\right)\left(\underline{\hat{\pi}} - \underline{\pi}\right)'\right] = \begin{bmatrix} f^{(-1)}\left(\hat{y}_{c_1}\right) & 0 & \cdots & 0 \\ 0 & f^{(-1)}\left(\hat{y}_{c_2}\right) & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & f^{(-1)}\left(\hat{y}_{c_q}\right) \end{bmatrix} V_c \begin{bmatrix} f^{(-1)}\left(\hat{y}_{c_1}\right) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f^{(-1)}\left(\hat{y}_{c_q}\right) \end{bmatrix} \quad (2.26)$$

2.1.3 RESPONSE TRANSFORMATION EQUATIONS

All equations presented thus far are valid regardless of the transformation employed to get from $\underline{p}$ to $\underline{y}$ provided all the quantities necessary for the estimation are defined. For the crop acreage proportion estimation problem, three transformations are potentially useful: the "identity," the "log," and the "logit." The formulas associated with these transformations and the peculiarities associated with their use are presented in sections 2.1.3.1, 2.1.3.2, and 2.1.3.3. For a more complete treatment of the development of these transformations, see reference 3.

2.1.3.1 The Identity Transformation

The identity transformation is:

$$f(p_i) = p_i \quad (2.27)$$

From the previous development, it follows that the W matrix associated with equation (2.27) has diagonal elements

$$w_{ii} = 1/\pi_i(1 - \pi_i) \quad (2.28)$$

In order to avoid numerical problems in the computation of $\hat{W}$, the SAS implementation replaces any $p_i < \epsilon_2$ by ϵ_2 and any $p_i > 1 - \epsilon_2$ by $1 - \epsilon_2$ when evaluating equation (2.11). The parameter ϵ_2 is specified by the user. (The substitution is also performed, if necessary, during each reweighting iteration.)

Since equation (2.27) is a linear function, it also follows that

$$
\begin{aligned}
\hat{\underline{a}} &= \hat{y}_c \\
\text{bias}\left[\hat{\underline{a}}\right] &= \underline{0}_q \\
\text{MMS}\left[\hat{\underline{a}}\right] &= \text{Var}\left[\hat{\underline{a}}\right] = V_c
\end{aligned} \tag{2.29}
$$

2.1.3.2 <u>The Log Transformation</u>

For the log transformation

$$f(p_i) = \ell n(p_i) \tag{2.30}$$

This implies the associated W matrix has diagonal entries

$$w_{ii} = \pi_i/(1 - \pi_i) \tag{2.31}$$

Also, since equation (2.30) is not linear, its application to equations (2.24) and (2.26) yields

$$
\begin{aligned}
\hat{\underline{a}} &= \left[\hat{a}_i\right] = \left[e^{\hat{y}_{c_i}}\right] \\
\text{bias}\left[\hat{\underline{a}}\right] &= (1/2)\begin{bmatrix} v_{c_{11}} & 0 & \cdots & 0 \\ 0 & v_{c_{22}} & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & v_{c_{qq}} \end{bmatrix}\left[e^{\hat{y}_{c_i}}\right]
\end{aligned} \tag{2.32}
$$

$$
\text{MMS}\left[\hat{\underline{a}}\right] \approx \begin{bmatrix} e^{\hat{y}_{c_1}} & 0 & \cdots & 0 \\ 0 & e^{\hat{y}_{c_2}} & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & e^{\hat{y}_{c_q}} \end{bmatrix} V_c \begin{bmatrix} e^{\hat{y}_{c_1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\hat{y}_{c_q}} \end{bmatrix}
$$

In the SAS implementation, several steps are taken to avoid numerical problems. First, the calculation of $\hat{W}$ is handled as it is in the case of the identity transformation. Second, since $\ln(p_i)$ is not defined for $p_i = 0$, a "working log" transformation is used which defines the elements of $\underline{y}$ as

$$y_i = \begin{cases} \ln(p_i) & \text{if } p_i > \varepsilon_1 \\ \ln\left(\frac{\varepsilon_1}{e}\right) + \frac{p_i}{\varepsilon_1} & \text{if } 0 \le p_i \le \varepsilon_1 \end{cases} \tag{2.33}$$

The effect of this working log transformation is illustrated in figure 2-1. (The parameter ε_1 must be specified by the user.) Third, to counteract the effect of using equation (2.33) rather than equation (2.27), the elements of $\hat{\underline{a}}$ are taken to be

$$\hat{a}_i = \begin{cases} e^{\hat{y}_{c_i}} & \text{if } \hat{y}_{c_i} > \ln(\varepsilon_1) \\ \left[\hat{y}_{c_i} - \ln\left(\frac{\varepsilon_1}{e}\right)\right]\varepsilon_1 & \text{if } \hat{y}_{c_i} \le \ln(\varepsilon_1) \end{cases} \tag{2.34}$$

which is the inverse of equation (2.23).

2.1.3.3 The Logit Transformation

For the logit transformation

$$f(p_i) = (1/2)\ln\left(\frac{p_i}{1 - p_i}\right) \tag{2.35}$$

and the associated W matrix has diagonal entries

$$w_{ii} = \pi_i(1 - \pi_i) \tag{2.36}$$

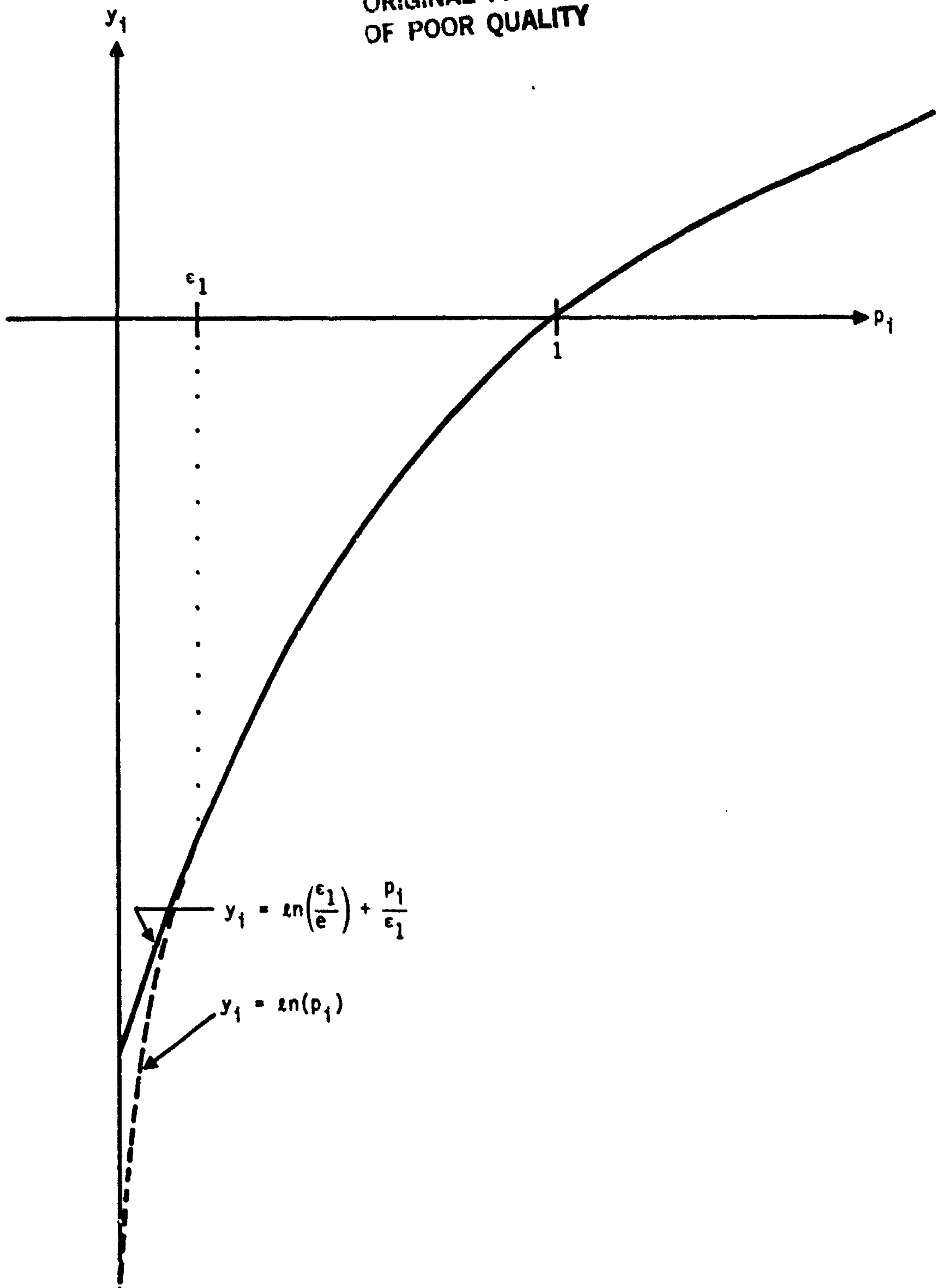

Figure 2-1.- Working log function.

Also,

$$\underline{\hat{a}} = \left[\hat{a}_i\right] = \left[e^{2\hat{y}_{c_i}} \Big/ \left(1 + e^{2\hat{y}_{c_i}}\right)\right]$$

$$\text{bias}\left[\underline{\hat{a}}\right] = \begin{bmatrix} v_{c_{11}} & 0 & \cdots & 0 \\ 0 & v_{c_{22}} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & v_{c_{qq}} \end{bmatrix} \left[\frac{2e^{\hat{y}_{c_i}}\left(1 - e^{2\hat{y}_{c_i}}\right)}{\left(1 + e^{2\hat{y}_{c_i}}\right)^3}\right] \tag{2.37}$$

$$\text{MSS}\left[\underline{\hat{a}}\right] = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & d_q \end{bmatrix} V_c \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & d_q \end{bmatrix}$$

where

$$d_i = 2e^{2\hat{y}_{c_i}} \Big/ \left(1 + e^{2\hat{y}_{c_i}}\right)^2$$

As before, the SAS implementation avoids difficulties in computing $\hat{W}$ by replacing values of p_i less than ε_2 with ε_2 (greater than $1 - \varepsilon_2$ with $1 - \varepsilon_2$). Also, since equation (2.35) is not defined for $P_i \varepsilon \{0, 1\}$, a working logit is used which defines the elements of $\underline{y}$ as

$$y_i = \begin{cases} \dfrac{\varepsilon_1}{\left\{\varepsilon_1 + (1 - \varepsilon_1)\exp\left[(1 - \varepsilon_1)^{-1}\right]\right\}} + \dfrac{p_i}{2\varepsilon(1 - \varepsilon_1)} & \text{if } 0 < p_i < \varepsilon_1 \\ (1/2)\ln\left(\dfrac{p_i}{1 - p_i}\right) & \text{if } \varepsilon_1 < p_i < 1 - \varepsilon_1 \\ \dfrac{-\varepsilon_1}{\left\{\varepsilon_1 + (1 - \varepsilon_1)\exp\left[(1 - \varepsilon_1)^{-1}\right]\right\}} + \dfrac{(p_i - 1)}{2\varepsilon_1(1 - \varepsilon)} & \text{if } 1 - \varepsilon_1 < p_i < 1 \end{cases} \tag{2.38}$$

The effect of this working transformation is illustrated in figure 2-2. In order to counteract the effect of using equation (2.38), the elements of $\underline{\hat{a}}$ are taken to be

$$\left\{\begin{array}{ll} \left(\hat{y}_{c_i} - \dfrac{\varepsilon_1}{\left\{\varepsilon_1 + (1 - \varepsilon_1)\exp\left[(1 - \varepsilon_1)^{-1}\right]\right\}}\right)2\varepsilon_1(1 - \varepsilon_1) & \text{if } \hat{y}_{c_i} < (1/2)\ln\dfrac{\varepsilon_1}{1 - \varepsilon_1} \\ \dfrac{e^{2\hat{y}_{c_i}}}{\left(1 + e^{2\hat{y}_{c_i}}\right)} & \text{if } (1/2)\ln\left(\dfrac{\varepsilon_1}{1 - \varepsilon_1}\right) < \hat{y}_{c_i} < (1/2)\ln\left(\dfrac{1 - \varepsilon_1}{\varepsilon_1}\right) \\ 1 + \left(\hat{y}_{c_i} + \dfrac{\varepsilon_1}{\left\{\varepsilon_1 + (1 - \varepsilon_1)\exp\left[(1 - \varepsilon_1)^{-1}\right]\right\}}\right)2\varepsilon_1(1 - \varepsilon_1) & \text{if } (1/2)\ln\left(\dfrac{1 - \varepsilon_1}{\varepsilon_1}\right) < \hat{y}_{c_i} \end{array}\right. \tag{2.39}$$

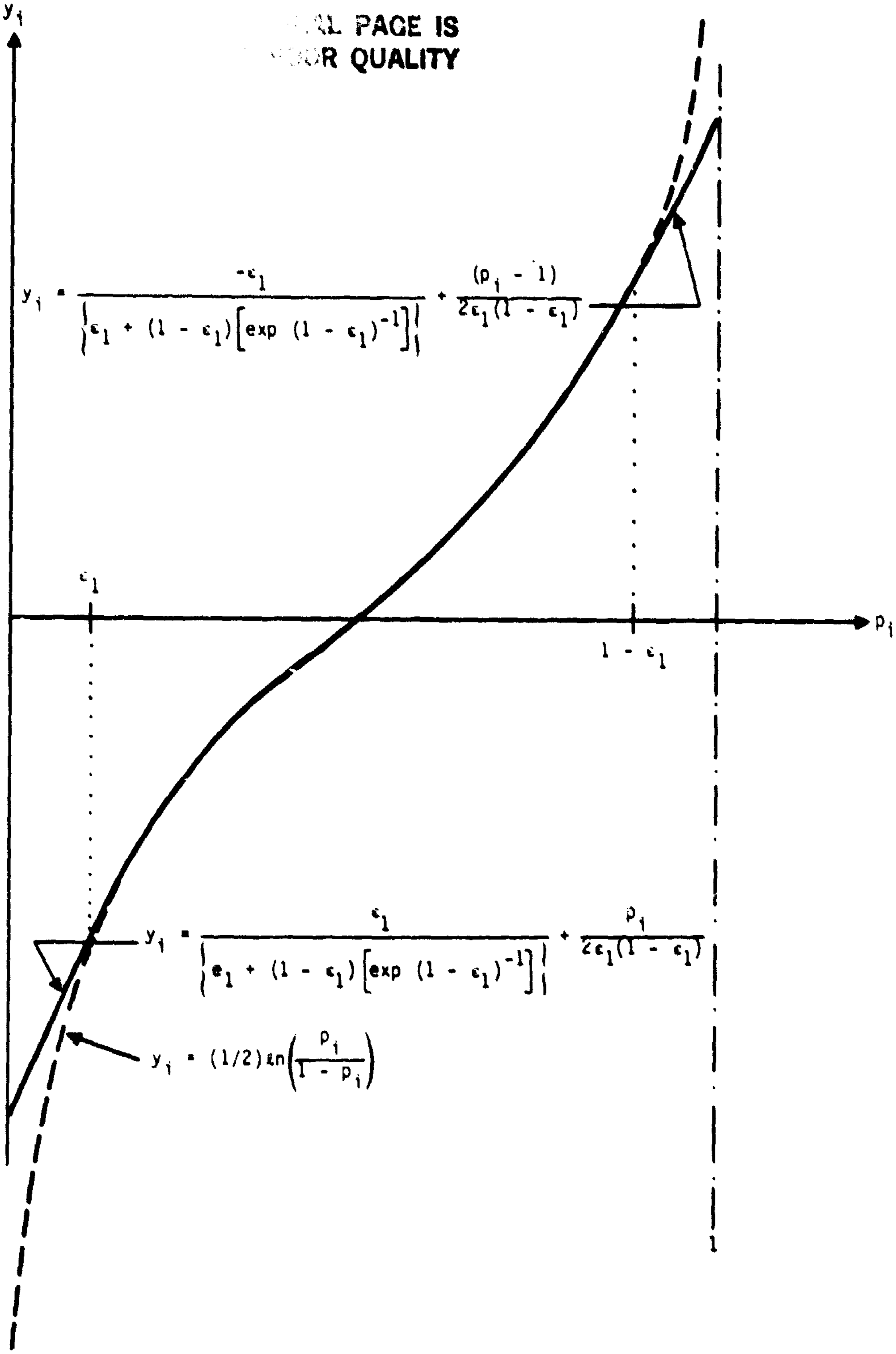

Figure 2-2.- Working logit function.

3. CROP ACREAGE PROPORTION ESTIMATION USING THE MIXED AOV MODEL

The development of the proposed methodology has been of a very general nature. Consequently, it has very broad applicability; and, as will be seen in the following paragraphs, the interpretation of the estimation results can vary widely depending on the data used and the design matrices (X,Z) specified. The remainder of this section will be devoted to presenting the analyses of specific models and data. All of the examples will be derived using subsets of the data shown in table 3-1.

Example 1: Suppose an estimate of the stratum's at-harvest crop acreage proportion is desired for each year having representative segment data in table 1. The model to be used is

$$y_{ijk} = \ell n(p_{ijk}) = \alpha_i + \delta_j + r_k + e_{ijk} \quad (3.1)$$

where

y_{ijk} = the transformed crop acreage proportion estimate for segment k at growth stage j in year i

α_i = the stratum's transformed crop acreage proportion in year i

δ_j = an adjustment for the tendency of the crop acreage proportion estimate for growth stage j to be different from the stratum's at-harvest crop acreage proportion

r_k = an adjustment for the tendency of the crop acreage proportion estimate for segment k to be different from the stratum's crop acreage proportion

e_{ijk} = all other unexplained effects

The year and growth stage effects are fixed; the segment effect is random.

TABLE 3-1.- DATA FOR CROP ACREAGE PROPORTION ESTIMATION EXAMPLES

Observation number	Segment crop acreage proportion estimate (a)	Identification of —			
		Crop year	Growth stage (b)	Segment	Substratum
1	0.279	1	2	1	1
2	.154	2	2	2	1
3	.149	3	2	2	1
4	.074	1	2	3	2
5	.073	2	2	3	2
6	.229	2	2	4	2
7	.212	3	2	4	2
8	.275	1	1	1	1
9	.152	2	1	2	1
10	.073	1	1	3	2
11	.069	2	1	3	2
12	.210	3	1	4	2

[a] As estimated from Landsat multispectral data

[b] 1 = Midseason; 2 = at-harvest

The matrix representation of equation (3.1) using the data shown in table 2-1 is

$$\begin{bmatrix} \ln(.279) \\ \ln(.154) \\ \ln(.149) \\ \ln(.074) \\ \ln(.073) \\ \ln(.229) \\ \ln(.212) \\ \ln(.275) \\ \ln(.152) \\ \ln(.073) \\ \ln(.069) \\ \ln(.210) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \delta_1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} + \begin{bmatrix} e_{121} \\ e_{222} \\ e_{322} \\ e_{123} \\ e_{223} \\ e_{224} \\ e_{324} \\ e_{111} \\ e_{212} \\ e_{113} \\ e_{213} \\ e_{314} \end{bmatrix} \quad (3.2)$$

Let C be defined as

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad (3.3)$$

Then $\hat{\underline{a}}$ is a 5-by-1 vector, and its elements are interpreted as follows:

a. $\hat{a}_1$ = an estimate of the stratum's at-harvest crop acreage proportion for year 1.

b. $\hat{a}_2$ = an estimate of the stratum's at-harvest crop acreage proportion for year 2.

c. $\hat{a}_3$ = an estimate of the stratum's at-harvest crop acreage proportion for year 3.

d. $\hat{a}_4$ = an estimate of the ratio of the stratum's at-harvest crop acreage proportions for years 1 and 2.

e. $\hat{a}_5$ = an estimate of the stratum's observable crop acreage proportion during growth stage 1 of year 3.

Note also that

a. $\hat{a}_3 - \hat{a}_2$ is an estimate of the change in the stratum's at-harvest crop acreage proportion from year 2 to year 3.

b. $\hat{a}_3 - \hat{a}_5$ is an estimate of the change in the stratum's observable crop acreage proportion from growth stage 1 to growth stage 2.

Since the SAS implementation provides the entire matrix of mean squared errors for $\hat{\underset{\sim}{a}}$, measures of dispersion can be calculated for all the estimators listed.

Example 2: Suppose the model is now

$$y_{ijk\ell} = \ell n(p_{ijk\ell}) = a_i + \delta_j + \rho_\ell + \delta\rho_{j\ell} + r_k + e_{ijk\ell} \tag{3.4}$$

where all terms are as defined for equation (3.1), with

a. ρ_ℓ = an adjustment for the tendency of crop acreage proportion estimates in substratum ℓ to be different from the stratum's crop acreage proportion.

b. $\delta\rho_{j\ell}$ = an adjustment due to the interaction of δ_j and ρ_ℓ.

Since X must be a full rank, a reasonable matrix representation of equation (3.4) using the data of table 3-1 is

$$
\begin{bmatrix} \ell n(.279) \\ \ell n(.154) \\ \ell n(.149) \\ \ell n(.074) \\ \ell n(.073) \\ \ell n(.229) \\ \ell n(.212) \\ \ell n(.275) \\ \ell n(.152) \\ \ell n(.073) \\ \ell n(.069) \\ \ell n(.210) \end{bmatrix}
=
\begin{bmatrix}
1 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & -1 & 0 \\
0 & 1 & 0 & 0 & -1 & 0 \\
0 & 1 & 0 & 0 & -1 & 0 \\
0 & 0 & 1 & 0 & -1 & 0 \\
1 & 0 & 0 & 1 & 1 & 1 \\
0 & 1 & 0 & 1 & 1 & 1 \\
1 & 0 & 0 & 1 & -1 & -1 \\
0 & 1 & 0 & 1 & -1 & -1 \\
0 & 0 & 1 & 1 & -1 & -1
\end{bmatrix}
\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \delta_1 \\ \rho_1 \\ \delta\rho_{11} \end{bmatrix}
+
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix}
+
\begin{bmatrix} e_{1211} \\ e_{2221} \\ e_{3221} \\ e_{1231} \\ e_{2232} \\ e_{2242} \\ e_{3242} \\ e_{1111} \\ e_{2121} \\ e_{1132} \\ e_{2132} \\ e_{3142} \end{bmatrix}
\quad (3.5)
$$

Let C be defined as

$$
C = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & -1 & 0
\end{bmatrix}
\quad (3.6)
$$

then $\hat{\underset{\sim}{a}}$ is interpreted as follows:

a. $\hat{a}_1$ = an estimate of the stratum's at-harvest crop acreage proportion for year 1.

b. $\hat{a}_2$ = an estimate of the at-harvest crop acreage proportion for substratum 1 in year 2.

c. $\hat{a}_3$ = an estimate of the at-harvest crop acreage proportion for substratum 2 in year 3.

Example 3: Suppose an estimate of the stratum's at-harvest crop acreage proportion for year 2 which uses only data from growth stage 2 in year 2 is desired.

An appropriate mixed AOV model is

$$y_k = (1/2)\ell n\left(\frac{p_i}{1 - p_i}\right) = \mu + r_k + e_k \tag{3.7}$$

The corresponding matrix representation of the data is

$$\begin{bmatrix} (1/2)\ell n\left(\frac{.154}{.846}\right) \\ (1/2)\ell n\left(\frac{.073}{.927}\right) \\ (1/2)\ell n\left(\frac{.229}{.771}\right) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \mu + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} \tag{3.8}$$

This representation, however, does not satisfy the requirement that $n > \text{rank}(X:Z) + 1$. The requirements for estimation can be met if the fixed-effect model

$$y_k = \mu + e_k \tag{3.9}$$

is considered instead.

The removal of the random effect from the model causes no computational difficulty — the estimation of γ need only be bypassed. In addition, if the true (unknown) value of γ is small relative to σ_ε^2 (also unknown), the models (3.7) and (3.9) are roughly equivalent in their effect. In the SAS implementation of the methodology, specification of a random effect matrix Z is optional.

The only appropriate matrix C for model (3.9) is C = [1]; hence $\hat{\underline{a}} = \hat{a}_1$, where $\hat{a}_1$ is an estimate of the stratum's at-harvest crop acreage proportion for year 2.

4. DOCUMENTATION OF THE SAS IMPLEMENTATION

The implementation of the methodology proposed in this paper has been developed using the PROC MATRIX feature of SAS (release 79.1) as implemented on the Earth Observations Division Laboratory System (EODLS) AS/3000 computer. It currently resides in accounts JSC1740 and DS40 under the program identifier GMYP/SAS. The designation of variable names within the program corresponds as closely as possible to the names and symbols used in this document. In addition, the program has a thorough internal documentation.

4.1 REQUIRED INPUTS

The GMYP/SAS program requires the following matrices as inputs: PARM, P, X, Z, and C.

The PARM matrix is a 1-by-4 array vector containing the values of operating parameters used by the algorithm. Specifically,

a. PARM(1,1) = the number of reweighting iterations that should be performed in arriving at the final estimate of W (usually 2 is a sufficient value.)

b. PARM(1,2) = a numerical designator indicating the transformation that should be employed (0, identity; 1, log; 2, logit).

c. PARM(1,3) = a specification for the value of ϵ_1 (usually .001).

d. PARM(1,4) = a specification for the value of ϵ_2 (usually .01).

The definitions of matrices P (i.e., $\underline{p}$), X, Z, and C are the same as those given in this document.

The current version of the program requires that the user initialize the matrices using the SAS assignment declarations within the program. Hence, the user must be familiar with the procedure for assigning values to matrices in PROC MATRIX. The code, however, may easily be modified to accept input at the time of execution.

If a model with no random effect is to be fitted, Z should be a column vector of zero's (i.e., $Z = \underline{0}_n$.) Also, the program verifies that the specified designs satisfy assumptions i and j of section 2. If not, processing terminates. If the specified design satisfies assumptions i and j, processing will continue even if the resulting model and estimates have no reasonable interpretation.

4.2 PROGRAM OUTPUTS

The GMYP/SAS program provides the user with a number of outputs, including the input data, intermediate parameter estimates, and final results. All outputs are generated using the PRINT option of PROC MATRIX. Hence, all numeric quantities are output in matrix format and are labeled by their interval variable names. Table 4-1 shows the variables for which values are output and the interpretation of these variables.

4.3 EXAMPLE RUNS

A sample input for examples 1, 2, and 3 of section 3 is shown in figure 4-1. Sample outputs for examples 1, 2, and 3 are given in figures 4-2, 4-3, and 4-4. The lines of code shown in figure 4-1 should be inserted into the SAS program at the location indicated on the program listing given in figure 4-5. A flow chart showing the functional flow of GYMP/SAS is shown in figure 4-6.

TABLE 4-1.- VARIABLES OUTPUT IN THE GMYP/SAS PROGRAM

Variable	Interpretation
PARM	Row vector of parameters
P	n-by-1 Vector of observed responses
X	n-by-t Fixed-effect design matrix
Z	n-by-s Random-effect design matrix
Y	Transformed-response vector
W	Final iteratively refined estimate of W
SIGMAE	Estimate of σ_ε^2
SIGMADE	Degrees of freedom upon which the estimate of σ_ε^2 is based
GAMMA	Estimate of γ
B	Estimate of $\underline{\beta} = \hat{\underline{\beta}}$
BVAR	Estimated V-C matrix for $\hat{\underline{\beta}}$
C	c-by-t contrast matrix
YA	Estimate of $C\underline{\beta} = \hat{\underline{y}}_c$
YAVAR	Estimated V-C matrix for $\hat{\underline{y}}_c$
A	Estimated value of $\underline{a} = \hat{\underline{a}}$
ABIAS	Estimated bias of $\hat{\underline{a}}$
AMSE	Estimated MMS for $\hat{\underline{a}}$
PDIF	Vector of prediction residuals between $\underline{p}$ and its corresponding mixed model estimate. (PDIF can be used to calculate measures of "goodness of fit" for the model and transformation used.)

```
* DATA FOR EXAMPLE1 IN IMPLEMENTATION DOCUMENT;
PARM=2 1 .001 .01;
P=.279/.154/.149/.074/.073/.229/.212/.275/.152/.073/.069/.210;
X=1 0 0 0/0 1 0 0/0 0 1 0/1 0 0 0/0 1 0 0/0 1 0 0/
  0 0 1 0/1 0 0 1/0 1 0 1/1 0 0 1/0 1 0 1/0 0 1 1;
Z=1 0 0 0/0 1 0 0/0 1 0 0/0 0 1 0/0 0 1 0/0 0 0 1/
  0 0 0 1/1 0 0 0/0 1 0 0/0 0 1 0/0 0 1 0/0 0 0 1;
C=1 0 0 0/0 1 0 0/0 0 1 0/1 -1 0 0/0 0 1 1;
```

a. Sample input, example 1.

```
* DATA FOR EXAMPLE2 IN IMPLEMENTATION DOCUMENT;
PARM=2 1 .001 .01;
P=.279/.154/.149/.074/.073/.229/.212/.275/.152/.073/.069/.210;
X=1 0 0 0  1  0/0 1 0 0  1  0/0 0 1 0  1  0/1 0 0 0 -1  0/
  0 1 0 0 -1  0/0 1 0 0 -1  0/0 0 1 0 -1  0/1 0 0 1  1  1/
  0 1 0 1  1  1/1 0 0 1 -1 -1/0 1 0 1 -1 -1/0 0 1 1 -1 -1;
Z=1 0 0 0/0 1 0 0/0 1 0 0/0 0 1 0/0 0 1 0/0 0 0 1/
  0 0 0 1/1 0 0 0/0 1 0 0/0 0 1 0/0 0 1 0/0 0 0 1;
C=1 0 0 0 0 0/0 1 0 0 1 0/0 0 1 0 -1 0;
```

b. Sample input, example 2.

```
* DATA FOR EXAPMLE3 IN IMPLEMENTATION DOCUMENT;
PARM=2 2 .001 .01;
P=.154/.073/.229;
X=1/1/1;
Z=0/0/0;
C=1;
```

c. Sample input, example 3.

Figure 4-1.- Sample input for examples 1, 2, and 3 of section 3.

ORIGINAL PAGE IS
OF POOR QUALITY

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M

PARM	COL1	COL2	COL3	COL4
ROW1	2	1	0.001	0.01

P	COL1
ROW1	0.279
ROW2	0.154
ROW3	0.149
ROW4	0.074
ROW5	0.073
ROW6	0.229
ROW7	0.212
ROW8	0.275
ROW9	0.152
ROW10	0.073
ROW11	0.069
ROW12	0.21

X	COL1	COL2	COL3	COL4
ROW1	1	0	0	0
ROW2	0	1	0	0
ROW3	0	0	1	0
ROW4	1	0	0	0
ROW5	0	1	0	0
ROW6	0	1	0	0
ROW7	0	0	1	0
ROW8	1	0	0	1
ROW9	0	1	0	1
ROW10	1	0	0	1
ROW11	0	1	0	1
ROW12	0	0	1	1

Z	COL1	COL2	COL3	COL4
ROW1	1	0	0	0
ROW2	0	1	0	0
ROW3	0	1	0	0
ROW4	0	0	1	0
ROW5	0	0	1	0
ROW6	0	0	0	1
ROW7	0	0	0	1
ROW8	1	0	0	0
ROW9	0	1	0	0
ROW10	0	0	1	0
ROW11	0	0	1	0
ROW12	0	0	0	1

Y	COL1
ROW1	-1.27654
ROW2	-1.8708
ROW3	-1.90381
ROW4	-2.60369
ROW5	-2.6173
ROW6	-1.47403
ROW7	-1.55117
ROW8	-1.29098
ROW9	-1.88387
ROW10	-2.6173
ROW11	-2.67365
ROW12	-1.56065

Figure 4-2.- Sample output for example 1.

ORIGINAL PAGE IS
OF POOR QUALITY

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M

W	COL1	COL2	COL3	COL4	COL5	COL6
ROW1	0.388701	0	0	0	0	0
ROW2	0	0.184369	0	0	0	0
ROW3	0	0	0.17213	0	0	0
ROW4	0	0	0	0.0802295	0	0
ROW5	0	0	0	0	0.0772788	0
ROW6	0	0	0	0	0	0.293761
ROW7	0	0	0	0	0	0
ROW8	0	0	0	0	0	0
ROW9	0	0	0	0	0	0
ROW10	0	0	0	0	0	0
ROW11	0	0	0	0	0	0
ROW12	0	0	0	0	0	0

W	COL7	COL8	COL9	COL10	COL11	COL12
ROW1	0	0	0	0	0	0
ROW2	0	0	0	0	0	0
ROW3	0	0	0	0	0	0
ROW4	0	0	0	0	0	0
ROW5	0	0	0	0	0	0
ROW6	0	0	0	0	0	0
ROW7	0.27259	0	0	0	0	0
ROW8	0	0.377576	0	0	0	0
ROW9	0	0	0.179849	0	0	0
ROW10	0	0	0	0.0784319	0	0
ROW11	0	0	0	0	0.0755519	0
ROW12	0	0	0	0	0	0.265423

SIGMAE	COL1
ROW1	0.000033604

SIGMADF	COL1
ROW1	5

GAMMA	COL1
ROW1	10992.7

H	COL1
ROW1	-1.78651
ROW2	-1.82137
ROW3	-1.8796
ROW4	-0.0210066

Figure 4-2.- Continued.

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M

BVAR	COL1	COL2	COL3	COL4
ROW1	0.0925477	0.0922792	0.0922554	-.000022708
ROW2	0.0922792	0.0924419	0.0924181	-.000022697
ROW3	0.0922554	0.0924181	0.0925046	-.000029671
ROW4	-.000022708	-.000022697	-.000029671	.0000594303

C	COL1	COL2	COL3	COL4
ROW1	1	0	0	0
ROW2	0	1	0	0
ROW3	0	0	1	0
ROW4	1	-1	0	0
ROW5	0	0	1	1

YA	COL1
ROW1	-1.78651
ROW2	-1.82137
ROW3	-1.8796
ROW4	0.0348533
ROW5	-1.90061

YAVAR	COL1	COL2	COL3	COL4	COL5
ROW1	0.0925477	0.0922792	0.0922554	0.000268564	0.0922327
ROW2	0.0922792	0.0924419	0.0924181	-.000162763	0.0923954
ROW3	0.0922554	0.0924181	0.0925046	-.000162715	0.0924749
ROW4	0.000268564	-.000162763	-.000162715	0.000431327	-.000162726
ROW5	0.0922327	0.0923954	0.0924749	-.000162726	0.0925046

A	COL1
ROW1	0.167543
ROW2	0.161805
ROW3	0.15265
ROW4	1.03547
ROW5	0.149477

ABIAS	COL1
ROW1	0.00775288
ROW2	0.00747876
ROW3	0.00706043
ROW4	0.000223312
ROW5	0.00691367

AMSE	COL1	COL2	COL3	COL4	COL5
ROW1	0.00259789	0.00250162	0.00235948	.0000465919	0.00230987
ROW2	0.00250162	0.00242019	0.00228268	-0.00002727	0.00223468
ROW3	0.00235948	0.00228268	0.00215556	-0.00002572	0.00211007
ROW4	.0000465919	-0.00002727	-0.00002572	0.000462466	-.000025187
ROW5	0.00230987	0.00223468	0.00211007	-.000025187	0.00206687

Figure 4-2.- Continued.

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M

PDIF	COL1
ROW1	0.111457
ROW2	-0.0078045
ROW3	-0.00365045
ROW4	-0.0935433
ROW5	-0.0888045
ROW6	0.0671955
ROW7	0.0593495
ROW8	0.110939
ROW9	-0.00644096
ROW10	-0.0910605
ROW11	-0.089441
ROW12	0.0605228

Figure 4-2.- Concluded.

STATISTICAL ANALYSIS SYSTEM

PARM	COL1	COL2	COL3	COL4
ROW1	2	1	0.001	0.01

P	COL1
ROW1	0.279
ROW2	0.154
ROW3	0.149
ROW4	0.074
ROW5	0.073
ROW6	0.229
ROW7	0.212
ROW8	0.275
ROW9	0.152
ROW10	0.073
ROW11	0.069
ROW12	0.21

X	COL1	COL2	COL3	COL4	COL5	COL6
ROW1	1	0	0	0	1	0
ROW2	0	1	0	0	1	0
ROW3	0	0	1	0	1	0
ROW4	1	0	0	0	-1	0
ROW5	0	1	0	0	-1	0
ROW6	0	1	0	0	-1	0
ROW7	0	0	1	0	-1	0
ROW8	1	0	0	1	1	1
ROW9	0	1	0	1	1	1
ROW10	1	0	0	1	-1	-1
ROW11	0	1	0	1	-1	-1
ROW12	0	0	1	1	-1	-1

Z	COL1	COL2	COL3	COL4
ROW1	1	0	0	0
ROW2	0	1	0	0
ROW3	0	1	0	0
ROW4	0	0	1	0
ROW5	0	0	1	0
ROW6	0	0	0	1
ROW7	0	0	0	1
ROW8	1	0	0	0
ROW9	0	1	0	0
ROW10	0	0	1	0
ROW11	0	0	1	0
ROW12	0	0	0	1

Y	COL1
ROW1	-1.27654
ROW2	-1.8708
ROW3	-1.90361
ROW4	-2.60369
ROW5	-2.6173
ROW6	.47403
ROW7	.55117
ROW8	1.29098
ROW9	-1.68387
ROW10	-2.6173
ROW11	-2.67365
ROW12	-1.56065

Figure 4-3.- Sample output for example 2.

ORIGINAL PAGE IS
OF POOR QUALITY

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M

W	COL1 COL7	COL2 COL8	COL3 COL9	COL4 COL10	COL5 COL11	COL6 COL12
ROW1	0.387995 0	0 0	0 0	0 0	0 0	0 0
ROW2	0 0	0.184067 0	0 0	0 0	0 0	0 0
ROW3	0 0	0 0	0.172161 0	0 0	0 0	0 0
ROW4	0 0	0 0	0 0	0.0803754 0	0 0	0 0
ROW5	0 0	0 0	0 0	0 0	0.0774223 0	0 0
ROW6	0 0	0 0	0 0	0 0	0 0	0.293795 0
ROW7	0 0.273154	0 0	0 0	0 0	0 0	0 0
ROW8	0 0	0 0.378282	0 0	0 0	0 0	0 0
ROW9	0 0	0 0	0 0.180122	0 0	0 0	0 0
ROW10	0 0	0 0	0 0	0 0.0782848	0 0	0 0
ROW11	0 0	0 0	0 0	0 0	0 0.0754138	0 0
ROW12	0 0	0 0	0 0	0 0	0 0	0 0.264619

SIGMAE	COL1
ROW1	.0000408873

SIGMADF	COL1
ROW1	4

GAMMA	COL1
ROW1	15874.5

B	COL1
ROW1	-1.78686
ROW2	-1.82155
ROW3	-1.87823
ROW4	-0.0213883
ROW5	0.23617
ROW6	0.0030828

Figure 4-3.- Continued.

ORIGINAL PAGE IS
OF POOR QUALITY

STATISTICAL ANALYSIS SYSTEM

BVAR	COL1	COL2	COL3	COL4	COL5	COL6
ROW1	0.162508	0.162182	0.162149	-.000026846	-.000043157	-6.3985E-06
ROW2	0.162182	0.16238	0.162347	-.000026837	.0000212425	-6.4068E-06
ROW3	0.162149	0.162347	0.162471	-.000040569	.0000164772	0.000036286
ROW4	-.000026846	-.000026837	-.000040569	.0000736227	.0000028548	-.000010475
ROW5	-.000043157	.0000212425	.0000164772	.0000028548	0.162306	-.000030988
ROW6	-6.3985E-06	-6.4068E-06	0.000036286	-.000010475	-.000030988	0.000084107

C	COL1	COL2	COL3	COL4	COL5	COL6
ROW1	1	0	0	0	0	0
ROW2	0	1	0	0	1	0
ROW3	0	0	1	0	-1	0

YA	COL1
ROW1	-1.78686
ROW2	-1.58538
ROW3	-2.1144

YAVAR	COL1	COL2	COL3
ROW1	0.162508	0.162138	0.162192
ROW2	0.162138	0.324728	.0000363368
ROW3	0.162192	.0000363368	0.324744

A	COL1
ROW1	0.167485
ROW2	0.20487
ROW3	0.120705

ABIAS	COL1
ROW1	0.0136088
ROW2	0.0332636
ROW3	0.0195992

AMSE	COL1	COL2	COL3
ROW1	0.00455855	0.0055634	0.00327893
ROW2	0.0055634	0.0136295	8.9857E-07
ROW3	0.00327893	8.9857E-07	0.00473145

PDIF	COL1
ROW1	0.066899
ROW2	-0.0508705
ROW3	-0.0445802
ROW4	-0.0582538
ROW5	-0.0547453
ROW6	0.101255
ROW7	0.0912947
ROW8	0.0657463
ROW9	-0.0491543
ROW10	-0.0560567
ROW11	-0.0556572
ROW12	0.0922126

Figure 4-3.- Concluded.

ORIGINAL PAGE IS
OF POOR QUALITY

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M

PARM	COL1	COL2	COL3	COL4
ROW1	2	2	0.001	0.01

P	COL1
ROW1	0.154
ROW2	0.073
ROW3	0.229

X	COL1
ROW1	1
ROW2	1
ROW3	1

Z	COL1
ROW1	0
ROW2	0
ROW3	0

Y	COL1
ROW1	-0.851783
ROW2	-1.27075
ROW3	-0.606983

W	COL1	COL2	COL3
ROW1	0.12002	0	0
ROW2	0	0.12002	0
ROW3	0	0	0.12002

SIGMAE	COL1
ROW1	0.0135231

SIGMADF	COL1
ROW1	2

GAMMA	COL1
ROW1	0

B	COL1
ROW1	-0.909838

BVAR	COL1
ROW1	0.0375578

Figure 4-4.- Sample output for example 3.

ORIGINAL PAGE IS
OF POOR QUALITY

STATISTICAL ANALYSIS SYSTEM

C	COL1
ROW1	1

YA	COL1
ROW1	-0.909838

YAVAR	COL1
ROW1	0.0375578

A	COL1
ROW1	0.139473

ABIAS	COL1
ROW1	0.00650058

AMSE	COL1
ROW1	0.00216405

PDIF	COL1
ROW1	0.0145272
ROW2	-0.0664728
ROW3	0.0895272

Figure 4-4.- Concluded.

```
FILE: GMYP     SAS      A    EODL / JOHNSON SPACE CENTER

* GENERAL MULTI-YEAR STRATUM PROPORTION ESTIMATION ROUTINE;
* WORKING VERSION - APRIL 22, 1981;
* PROGRAMMED BY:  T.BAKER / LOCKHEED-EMSCO (245169);
*;
*;
* SAS PROGRAM TO PERFORM MULTI-YEAR ANALYSIS OF SEGMENT PROPORTION ESTIMATES;
* VIA ITERATIVELY REWEIGHTED LEAST SQUARES ANALYSIS OF THE MIXED AOV MODEL:;
*;
*      Y(P) = X*B + Z*R + E;
*;
*      WHERE P = THE (RESPONSE) VECTOR OF SEGMENT PROPORTION ESTIMATES;
*            Y = A TRANSFORMATION ON THE RESPONSE VECTOR;
*            X = THE DESIGN MATRIX FOR THE FIXED EFFECTS;
*            B = THE VECTOR OF FIXED EFFECTS;
*            Z = THE DESIGN MATRIX FOR THE RANDOM EFFECTS;
*            R = THE VECTOR OF RANDOM EFFECTS;
*            E = THE VECTOR OF 'UNEXPLAINED' ERRORS;
*;
*      MODEL ASSUMPTIONS:;
*            1) P IS A RANDOM VECTOR WITH MEAN PI AND VARIANCE;
*               PROPORTIONAL TO PI*(1-PI);
*            2) R IS A RANDOM VECTOR WITH MEAN ZERO AND DIAGONAL V-C MATRIX;
*               (DIAGONAL ELEMENTS = GAMMA*SIGMAE);
*            3) E IS A RANDOM VECTOR WITH MEAN ZERO AND DIAGONAL V-C MATRIX;
*               (DIAGONAL ELEMENTS = SIGMAE*VAR(Y(P)) = INV(W)*SIGMAE);
*            4) R AND E ARE INDEPENDENT;
*            5) THE MATRICES X,Z ARE OF FULL COLUMN RANK;
*               (EXCEPTION: Z SHOULD BE A COLUMN VECTOR OF ZEROES IF;
*                A MODEL WITH NO RANDOM EFFECT (R) IS TO BE FITTED);
*            6) Y(P) = THE IDENTITY, LOG, OR LOGIT TRANSFORMATION;
*            7) ADDITIONAL ASSUMPTIONS ON THE EFFECTS MAY BE;
*               INCORPORATED INTO THE DESIGN MATRICES: X,Z;
*            8) THE NUMBER OF OBSERVATIONS IS AT LEAST RANK(X:Z)+1;
*;
*      REQUIRED INPUTS: PARM, P, X, Z, C;
*;
*      OUTPUTS: PARM, P, X, Z, Y, W, SIGMAE, SIGMADF, GAMMA, B, BVAR;
*               C, YA, YAVAR, A, ABIAS, AMSE, PDIF;
*;
*      INTERNAL VARIABLE NAMES AND USAGES:;
*            PARM = A ROW VECTOR OF ALGORITHM PARAMETERS;
*               PARM(1,1)=NUMBER OF REWEIGHTING ITERATIONS (USUALLY 2);
*               PARM(1,2)=TRANSFORMATION TYPE (0-IDENTITY,1-LOG,2-LOGIT);
*               PARM(1,3)=TOLERANCE LIMIT FOR 'WORKING' FUNCTIONS (USUALLY .001);
*               PARM(1,4)=TOLERANCE LIMIT FOR WEIGHT CALCULATIONS (USUALLY .01);
*            P = THE COLUMN VECTOR OF RESPONSES;
*            X = THE FIXED EFFECTS DESIGN MATRIX;
*            Z = THE RANDOM EFFECTS DESIGN MATRIX;
*            C = A MATRIX WHOSE ROWS CONSIST OF FIXED EFFECTS DESIGN SETTINGS;
*               FOR WHICH PREDICTED VALUES OF TRANSFORMED-RESPONSE (YA) AND;
*               RESPONSE (A) ARE DESIRED;
*;
*            Y = THE TRANSFORMED RESPONSE VECTOR;
*            W = A WORK MATRIX WHOSE FINAL VALUE IS THE LAST UPDATED WEIGHTING;
*               MATRIX FOR THE MODEL DURING STEP ONE;
*            SIGMAE = THE ESTIMATED TRANSFORMED-ERROR VARIANCE = VAR(SQRT(W)*E);
*            SIGMADF = THE ERROR VARIANCE DEGREES OF FREEDOM;
*            GAMMA = THE ESTIMATED RANDOM-EFFECTS TO TRANSFORMED-ERROR;
*               VARIANCE RATIO;
*            B = THE VECTOR OF FIXED EFFECT ESTIMATES;
*            BVAR = THE ESTIMATED VARIANCE-COVARIANCE MATRIX FOR B;
*            YA = C*B = THE PREDICTED TRANSFORMED RESPONSES (Y-VALUES) FOR;
*               THE FIXED EFFECT DESIGN SETTINGS SPECIFIED IN C;
*            YAVAR = THE ESTIMATED VARIANCE-COVARIANCE MATRIX FOR THE PREDICTED;
*               VALUES IN YA;
*            A = THE VECTOR OF INVERSED-TRANSFORMED YA-VALUES;
*              = THE PREDICTED RESPONSES FOR THE DESIGN SETTINGS IN C;
*            ABIAS = A VECTOR OF APPROXIMATE BIAS ESTIMATES FOR A;
*            AMSE = THE ESTIMATED VARIANCE-COVARIANCE MATRIX FOR THE PREDICTED;
*               VALUES IN A;
*            PDIF = P - PHAT = THE FINAL VECTOR OF PREDICTION ERRORS;
*;
*            NR = NUMBER OF ROWS(OBSERVATIONS) IN X, P, Z, U;
*            NRC = NUMBER OF ROWS IN C;
*            NCX = NUMBER OF COLUMNS IN X;
*            NCZ = NUMBER OF COLUMNS IN Z;
*            XZ = THE CONCATENATION OF X,Z = (X:Z);
*            RANKXZ = RANK OF (X:Z)'*(X:Z);
*            PHAT = A WORK VECTOR WHOSE FINAL VALUE IS THE PREDICTED RESPONSE;
```

Figure 4-5.- Program listing for program GMYP/SAS.

ORIGINAL PAGE IS
OF POOR QUALITY


```
*              CORRESPONDING TO P;
*            YHAT = THE PREDICTED TRANSFORMED-RESPONSE VECTOR CORRESPONDING TO P;
*            V = THE WEIGHTING MATRIX FOR THE MODEL DURING STEP TWO;
*            MISCELLANEOUS CONSTANTS AND INTERMEDIATE RESULTS:;
*              ZXWXZI, XWXI, XVXI, PW, Y1, Y0, X0, B0, SLOPE, TEMP, SSE, SSB;
*;                                                                        -;
* -------------------------------------------------------------------------;
PROC MATRIX;
*;
* INPUT ALGORITHM PARAMETERS IN THE ROW VECTOR, PARM;;
* INPUT THE DATA: P, X, Z, C;
*;
* INSERT INPUT FILE HERE
* DATA FOR EXAMPLE; IN IMPLEMENTATION DOCUMENT;
PARM=2 1 .001 .01;
P=.279/.154/.149/.074/.073/.229/.212/.275/.152/.073/.069/.210;
X=1 0 0 0/0 1 0 0/0 0 1 0/1 0 0 0/0 1 0 0/0 1 0 0/
  0 0 1 0/1 0 0 1/0 1 0 1/1 0 0 1/0 1 0 1/0 0 1 1;
Z=1 0 0 0/0 1 0 0/0 1 0 0/0 0 1 0/0 0 1 0/0 0 0 1;
  0 0 0 1/1 0 0 0/0 1 0 0/0 1 0 0/0 0 1 0/0 0 1 1;
C=1 0 0 0/0 1 0 0/0 0 1 0/1 -1 0 0/0 0 1 1;
*;
* BEGIN STEP ONE: PERFORM ITERATIVELY REWEIGHTED LEAST SQUARES ANALYSIS;
*     OF THE MODEL - THEN CALCULATE SIGMAE AND GAMMA;
*;
* INITIALIZE FOR FIRST ITERATION AND PRINT DATA,PARAMETERS;
PRINT PARM P;
I=0; NR=NROW(X); NRC=NROW(C); NCX=NCOL(X); PHAT=P; Y=P;
IF ANY(Z) THEN NCZ=NCOL(Z); ELSE NCZ=0;
IF ANY(Z) THEN XZ=X||Z; ELSE XZ=X;
PRINT X Z;
*;
* CHECK FOR SINGULAR DESIGN MATRICES;
IF DET(X'*X)<.00001 THEN DO;
 NOTE 'THE DESIGN MATRIX X IS SINGULAR - PROCESSING STOPS'; STOP; END;
IF ANY(Z) THEN IF DET(Z'*Z)<.00001 THEN DO;
 NOTE 'THE DESIGN MATRIX Z IS SINGULAR - PROCESSING STOPS'; STOP; END;
RANKXZ=SUM(EIGVAL(XZ'*XZ)>.00001);
IF NR<=RANKXZ THEN DO;
 NOTE 'INSUFFICIENT DATA TO ESTIMATE MODEL PARAMETERS - PROCESSING STOPS';
 STOP; END;
*;
* PERFORM THE SELECTED TRANSFORMATION ON THE RESPONSE, P;
* COMPUTE THE INITIAL WEIGHTING MATRIX;
IF PARM(1,2)=0 THEN LINK WORKY0;
IF PARM(1,2)=1 THEN LINK WORKY1;
IF PARM(1,2)=2 THEN LINK WORKY2;
PRINT Y;
*;
* ITERATE TO DETERMINE THE REFINED WEIGHT MATRIX, W;
DO I=1 TO PARM(1,1);
 YHAT=XZ*GINV(XZ'*W*XZ)*XZ'*W*Y;
 IF PARM(1,2)=0 THEN LINK PHATY0;
 IF PARM(1,2)=1 THEN LINK PHATY1;
 IF PARM(1,2)=2 THEN LINK PHATY2;
END;
*;
* ESTIMATE SIGMAE AND GAMMA VIA HENDERSON'S METHOD THREE (REF: SEARLE);
XWXI=INV(X'*W*X);
ZXWXZI=GINV(XZ'*W*XZ);
SSE=Y'*W*(I.(NR)-XZ*ZXWXZI*XZ'*W)*Y;
SSB=Y'*W*((XZ*ZXWXZI*XZ')-(X*XWXI*X'))*W*Y;
SIGMADF=NR-RANKXZ;
SIGMAE=SSE#/SIGMADF;
IF ¬ANY(Z) THEN GAMMA=0;
ELSE GAMMA=((SSB#/SIGMAE)-RANKXZ+NCZ)#/TRACE(Z'*W*(I.(NR)-X*XWXI*X'*W)*Z);
*;
* STOP IF GAMMA<0;
IF GAMMA<0 THEN DO;
 NOTE THE ESTIMATE OF GAMMA IS NEGATIVE - PROCESSING STOPS; STOP; END;
PRINT W;
PRINT SIGMAE SIGMADF GAMMA;
*;
*;
* BEGIN STEP TWO: CALCULATE PREDICTIONS AND VARIANCES;
*;
* CALCULATE THE FINAL FIXED EFFECT ESTIMATES AND VARIANCES;
* CALCULATE THE PREDICTED TRANSFORMED-RESPONSES FOR THE DESIGN MATRIX C;
IF ¬ANY(Z) THEN V=W; ELSE V=INV(INV(W)+(Z*Z'#GAMMA));
```

Figure 4-5.- Continued.

```
XVXI=INV(X'*V*X);
B=XVXI*X'*V*Y;
BVAR=XVXI#SIGMAE;
YHAT=X*B;
YA=C*B; YAVAR=C*BVAR*C';
PRINT B BVAR C YA YAVAR;
*;
* CALCULATE RESIDUALS AND INVERSE-TRANSFORMED PARAMETER ESTIMATES;
IF PARM(1,2)=0 THEN LINK INVY0;
IF PARM(1,2)=1 THEN LINK INVY1;
IF PARM(1,2)=2 THEN LINK INVY2;
PRINT A ABIAS AMSE;
PDIF=P-PHAT; PRINT PDIF;
RETURN;
*;
* ------------------------------------------------------------------------;
*;
* FORMULAS ASSOCIATED WITH Y0, THE IDENTITY TRANSFORMATION;
WORKY0: * DEFINE THE 'WORKING IDENTITY' FUNCTION;
 Y=P; YHAT=Y;
PHATY0: * DEFINE THE INVERSE IDENTITY TRANSFORMATION ON THE RESPONSE;
 PHAT=YHAT;
 * DEFINE THE WEIGHTING MATRIX;
 PW=(PHAT<>PARM(1,4))><(1-PARM(1,4));
 W=DIAG(J,(NR,1,1)#/(PW#(J,(NR,1,1)-PW)));
 GOTO CONTINUE;
INVY0: * DEFINE THE INVERSE IDENTITY TRANSFORMATION ON YA=C*B;
 A=YA;
 ABIAS=J,(NRC,1,0);
 AMSE=YAVAR;
 PHAT=YHAT;
 GOTO CONTINUE;
*;
* FORMULAS ASSOCIATED WITH Y1, THE LOG TRANSFORMATION;
WORKY1: * DEFINE THE 'WORKING LOG' FUNCTION;
 X0=PARM(1,3); Y0=LOG(X0#/EXP(1)); SLOPE=1#/X0;
 DO J=1 TO NR;
 IF P(J,1)<=X0 THEN Y(J,1)=Y0+(P(J,1)#SLOPE);
 ELSE Y(J,1)=LOG(P(J,1));
 END;
 YHAT=Y;
PHATY1: * DEFINE THE INVERSE LOG TRANSFORMATION ON THE RESPONSE;
 PHAT=EXP(YHAT);
 * DEFINE THE WEIGHTING MATRIX;
 PW=(PHAT<>PARM(1,4))><(1-PARM(1,4));
 W=DIAG(PW#/(J,(NR,1,1)-PW));
 GOTO CONTINUE;
INVY1: * DEFINE THE INVERSE LOG TRANSFORMATION ON YA=C*B;
 A=YA; Y1=LOG(X0);
 DO J=1 TO NRC;
  B0=YA(J,1);
  IF B0<Y0 THEN A(J,1)=0.;
   ELSE IF B0<Y1 THEN A(J,1)=(B0-Y0)#/SLOPE;
   ELSE A(J,1)=EXP(B0);
  END;
 TEMP=EXP(YA);
 ABIAS=.5#VECDIAG(YAVAR)#TEMP;
 AMSE=DIAG(TEMP)*YAVAR*DIAG(TEMP);
 PHAT=EXP(YHAT);
 GOTO CONTINUE;
*;
* FORMULAS ASSOCIATED WITH Y2, THE LOGIT TRANSFORMATION;
WORKY2: * DEFINE THE 'WORKING LOGIT' FUNCTION;
 X0=PARM(1,3);
 Y0=X0#/(X0+((1-X0)#EXP(1#/(1-X0))));
 Y0=.5#LOG(Y0#/(1-Y0));
 SLOPE=1#/(2#X0#(1-X0));
 DO J=1 TO NR;
 IF P(J,1)<=X0 THEN Y(J,1)=Y0+(P(J,1)#SLOPE);
 ELSE IF P(J,1)<1-X0 THEN Y(J,1)=.5#LOG(P(J,1)#/(1-P(J,1)));
 ELSE Y(J,1)=((P(J,1)-1)#SLOPE)-Y0;
 END;
 YHAT=Y;
PHATY2: * DEFINE THE INVERSE LOGIT TRANSFORMATION ON THE RESPONSE;
 PHAT=EXP(2#YHAT)#/(J,(NR,1,1)+EXP(2#YHAT));
 * DEFINE THE WEIGHTING MATRIX;
 PW=(PHAT<>PARM(1,4))><(1-PARM(1,4));
 W=DIAG(PW#(J,(NR,1,1)-PW));
 GOTO CONTINUE;
```

Figure 4-5.- Continued.

ORIGINAL PAGE IS
OF POOR QUALITY

```
FILE: GMYP     SAS      A    EODL / JOHNSON SPACE CENTER

INVY2: * DEFINE THE INVERSE LOGIT TRANSFORMATION ON YA=C*B;
 A=YA; Y1=.5#LOG(X0#/(1-X0));
 DO J=1 TO NRC;
  B0=YA(J,1);
  IF B0<Y0 THEN A(J,1)=0.;
   ELSE IF B0<Y1 THEN A(J,1)=(B0-Y0)#/SLOPE;
   ELSE IF B0>-Y0 THEN A(J,1)=1.;
   ELSE IF B0>-Y1 THEN A(J,1)=(B0+Y0)#/SLOPE;
   ELSE A(J,1)=EXP(2#B0)#/(1+EXP(2#B0));
  END;
 TEMP=EXP(2#YA);
 ABIAS=VECDIAG(YAVAR)#2#TEMP#(J.(NRC,1,1)-TEMP)#/((J.(NRC,1,1)+TEMP)##3);
 TEMP=2#TEMP#/(J.(NRC,1,1)+TEMP)##2;
 AMSE=DIAG(TEMP)*YAVAR*DIAG(TEMP);
 PHAT=EXP(2#YHAT)#/(J.(NR,1,1)+EXP(2#YHAT));
CONTINUE: RETURN;
```

Figure 4-5.- Concluded.

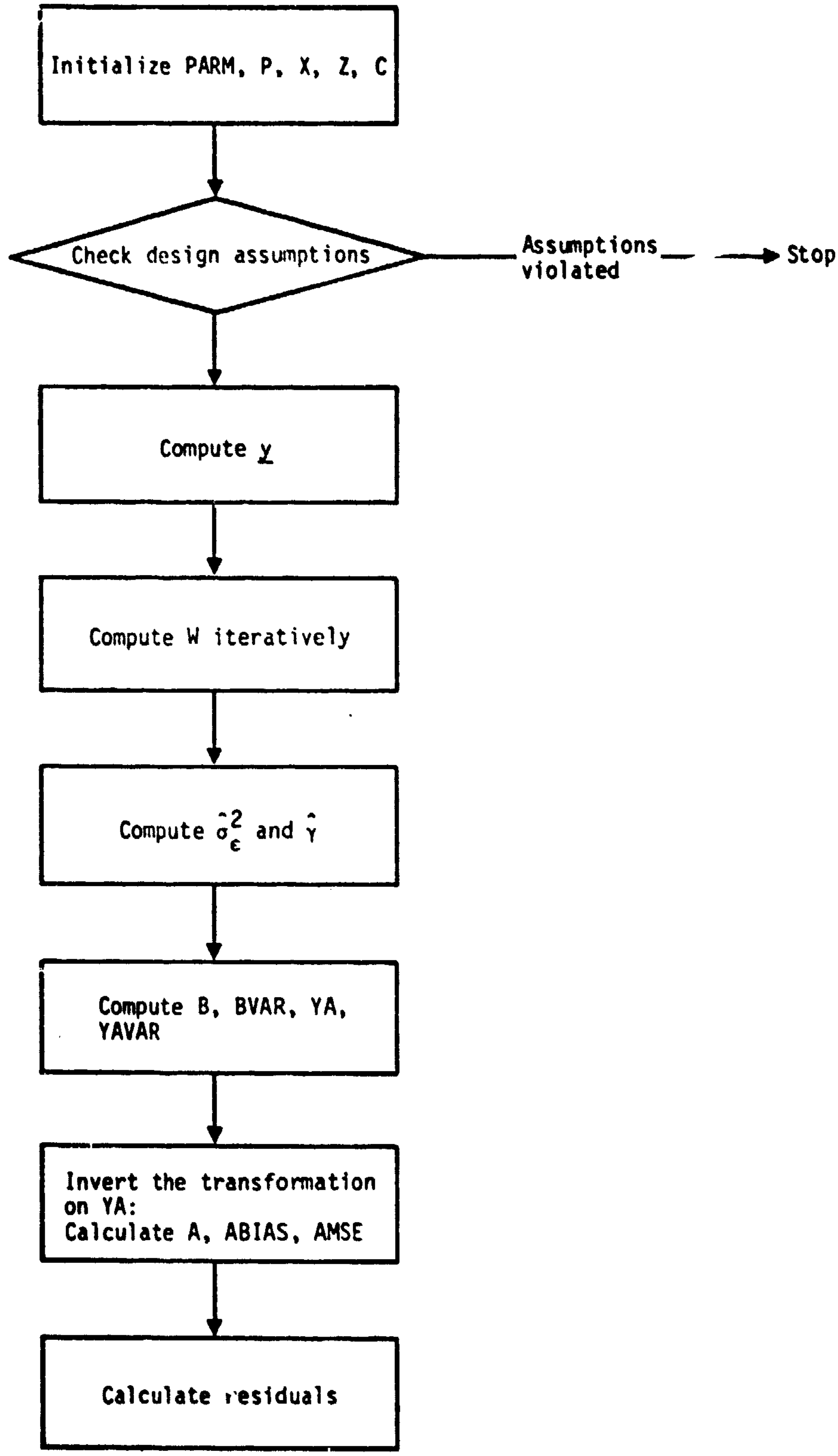

Figure 4-6.- Functional flow chart for program GMYP/SAS.

5. REFERENCES

1. Hartley, H. O.; Freund, R. J.; and Lee, T.: Gains of Precision Achievable by Multi-Year Estimation. Technical Report S- Institute of Statistics, Texas A&M Univ. (College Station, Texas), May 1979.

2. Searle, S. R.: Linear Models. John Wiley and Sons, chapter 10, 1971.

3. Sielken, R. L. Jr.; and Dahn, P. F: Multi-Year Estimation of the At-harvest Crop Acreage Proportion: Methodology and Implementation. Technical Report 20, Institute of Statistics, Texas A&M Univ. (College Station, Texas), March 1981.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1982—569-011/724