77

BARD

FINAL REPORT

PROJECT No. I-217-80

**Estimation of an Agricultural Production
Function from Data Obtained from
Complex Sample Surveys**

G. Nathan, G. Bental, W.A. Fuller

1983

Jerusalem, July 1983

BARD
P.O. Box 6
Bet Dagan, ISRAEL

BARD Project No. I-217-80

Title of Research Project: Estimation of an Agricultural Production
Function from Data Obtained from Complex
Sample Surveys.

Name of Investigators:

Principal Investigator: Nathan, Gad

Cooperating Investigators: Fuller, Wayne A. and Bental, Gad

Name and Address of
Affiliated Institutions:

Grantee Institution: Institute of Farm Income Research
6 Sderot Hakhashmonaim
Tel Aviv, Israel

Cooperating Institution: Central Bureau of Statistics
Hakirya Romema, P.O.B. 13015
Jerusalem, Israel

Project's starting date: November 1, 1980

Type of report: Final

Signatures:

Principal Investigator

Institution's Authorizing Official

Gad Nathan

M. Siron

Abstract

The main problem of regression analysis of data coming from a complex survey design is that possibly different regression functions fit different groups in the finite population. It is assumed that the different regression functions are partially or fully exchangeable and one estimator is sought. An estimator minimizing the expected squared error loss over all the finite population is obtained. Its optimality properties are examined as are several special cases. Covariance components in the models involved are estimated by two methods, one is a component-by-component approach, the other is the EM algorithm. The methods are used in order to analyze data from a profitability survey of the grapefruit crop. The analysis indicates that productivity and profitability, are explained mainly by direct input to the orchard and by the age of the orchard. Different regions were found to have different intercepts but equal slopes. The covariance estimation methods applied to the data set were used to obtain estimators of the ultimate model parameters.

Table of Contents

Abstract	i
The Effect of Sampling Design on Regression Analysis in Finite Population Sampling	1
1. Introduction and Literature Review	1
2. The Model - Assumptions and Estimation	3
3. A Special Case: $A_i = I$	8
4. Variance Estimation	13
5. The Analysis of the Data	20
6. Conclusions	39
7. Bibliography	41
Description and results of cooperation	43
Computer programs	44

The Effect of Sampling Design on Regression Analysis in Finite Population Sampling

1. Introduction and Literature Review

In studies employing regression techniques we often find that the population consists of different groups for which the relationships between the variables may differ. Conclusions may be sought on a macro level and if the usual assumptions of the regression model do not hold, the results will be erroneous or strongly biased. For example, suppose a production function in agriculture is constructed. The different geographical regions are obvious different groups, but we may still want one overall function estimating effects for the whole country. Moreover, even if the investigators are interested in different estimators for different groups, practitioners frequently find that they have no or insufficient data on the separate groups (e.g. small samples in strata or cluster sampling) to obtain good estimates of all the parameters involved, e.g. the vector of coefficients for every group.

The units may be sampled by a known and well defined probability sampling scheme based on the grouping in the population (e.g. stratification or clustering). There may be a relation between the sampling scheme and the models involved. If so, this should be taken into account in the analysis.

One of the ways to overcome problems of grouping in the population and the need for a unified approach to the analysis is to modify ordinary least squares analysis. This is done by dividing each product or square of the observed values of the explained and explanatory variables by the unit's associated inclusion probability. This provides a model-free design-unbiased estimation of the overall finite population regression coefficients - see Kish and Frankel (1974). There can be two reasons to use the inclusion probabilities even under a model-based approach. One is that the variance may be closely related to the inclusion probabilities, Klein (1953, p. 309), so that the

weighting provides a more efficient estimator than OLS, closer to the generalized least square estimator. The other possible reason for the use of inclusion probabilities is because of robustness considerations. In other words, an appropriate weighting of the sample observations can protect against certain deviations from the underlying model. Work in this direction is reported by Klein and Morgan (1951), Klein (1953 p. 305), Hu and Stromsdorfer (1970), Kish and Frankel (1974), Holt, Smith & Winter (1980), Nathan and Holt (1980). Hu and Stromsdorfer consider the case of stratified samples with differential sampling fractions in various strata. They propose a method which takes into account the design of the sample, but still assumes the same model and values of parameters in the different strata. Lindley & Smith (1972) introduce the idea of exchangeability and extend it to a multiple regression case. The idea is that improved estimates might be obtained for any one group by combining the data from all groups sampled whenever exchangeability between regressions is a sensible assumption. This idea is further extended here and applied to a situation of complex survey design. Fay & Herriot (1979) also allow different coefficients for different groups and use James-Stein procedures to estimate them. The estimates presented here are another generalization of similar ideas with slightly different estimators.

When applying exchangeability to a situation of complex survey design, the problem of variance-covariance estimation is encountered. Because of the complex survey design, standard variance estimation procedures, such as based on O.L.S. or maximum likelihood become complicated. The literature on covariance components models is extensive and will not be reviewed here. One approach we have used utilizes the EM algorithm and is based on Dempster, Rubin and Tsutakawa (1981). Another approach used is based on estimating each component separately. This approach is an extension of the one-parameter case presented in Pfeiffermann & Nathan (1981).

2. The Model - Assumptions and Estimation

Suppose there is a finite population U of M identifiable units, which are divided, into N groups, the i -th of which has M_i units. It is assumed that the same set of explanatory variables explains the dependent variable for all groups in the population. This assumption is only technical and does not impose any loss of generality, since the regression coefficients are not required to be non-zero for all variables in the set in every group. However, different vectors of regressions coefficients $\underline{\beta}_i$ are allowed in each group.

Denote by Y_{ij} the dependent variable for unit j in the i -th group and by $X_{ij} = (1, X_{ij1}, \dots, X_{ijk})'$ the corresponding values of the explanatory variables. Then the basic regression model is

$$Y_{ij} = \beta_{i0} + \beta_{i1} X_{ij1} + \dots + \beta_{ik} X_{ijk} + \epsilon_{ij}. \quad (1)$$

Denote $\underline{Y}_i = (Y_{i1}, \dots, Y_{iM_i})'$, $\underline{Y} = (\underline{Y}_1', \dots, \underline{Y}_N')$, $\underline{\beta}_i = (\beta_{i0}, \dots, \beta_{ik})$, $\tilde{\underline{\beta}} = (\underline{\beta}_1', \dots, \underline{\beta}_N')'$ and let X be the $M \times N(k+1)$ block diagonal matrix the i -th block of which is $(X_{i1}, \dots, X_{iM_i})'$ of order $M_i \times (k+1)$. Then (1) can be written $E(\underline{Y}) = X\tilde{\underline{\beta}}$. Furthermore, let $A = (A_1', A_2', \dots, A_N')'$ be an $N(k+1) \times r$ matrix of known constants, $\underline{\beta} = (\beta_1, \dots, \beta_r)'$ a parameter vector and Δ a $(k+1) \cdot (k+1)$ matrix.

Then the following is assumed:

1. The X -values are fixed.
2. $E(\epsilon_{ij}) = 0$; $\text{Var}(\epsilon_{ij}) = \sigma^2$; $\text{Cov}(\epsilon_{ij}, \epsilon_{kl}) = 0$ for $(i,j) \neq (k,l)$.

3. $\tilde{\underline{\beta}} = A\underline{\beta} + \underline{v}$; $E(\underline{v}) = 0$ $\text{Var}(\underline{v}) = \text{diag}(\Delta, \Delta, \dots, \Delta)$.
4. The sampling scheme is independent of the realizations $\underline{\beta}_i$ and of the residuals ϵ_{ij} .
5. X and A are full rank matrices.

Suppose that $n \leq N$ groups are sampled and $m_i \leq n_i$ units are sampled in the i -th group. One can view the model as

$$\tilde{\underline{Y}} = \tilde{X}(\tilde{\underline{\beta}}'; \underline{\beta}')' + (\underline{\epsilon}'_s, \underline{v}')' = \tilde{X}\underline{\beta}_0 + (\underline{\epsilon}'_s, \underline{v}')' \quad (2)$$

Where $\tilde{\underline{Y}} = (\underline{Y}'_s; 0_{N(k+1)})'$, $\underline{Y}_s$ and $\underline{\epsilon}_s$ are the vectors of sampled values of $\underline{Y}$ and $\underline{\epsilon}$ respectively and $\tilde{X}$ is the appropriate $\sum_{m_i} \times [N(k+1) + r]$ design matrix.

The purpose of the analysis is to predict values of the dependent variable Y with minimum loss. A formal application of Aitken's (Duncan and Horn (1972)) generalized least squares approach (even though the assumption of independence between coefficients and errors does not hold) gives estimators both for $\underline{\beta}$ and for $\underline{\beta}_i$, the regression coefficients of each group. Denoting the $m_i \times (k+1)$ block of the design matrix for the sample of the i -th group by $X_i = (X_{i1}', \dots, X_{im_i}')$, the estimators are:

$$\underline{\beta}(a) = \left\{ \sum_{i=1}^n A_i' X_i' (\sigma^2 I_{m_i} + X_i \Delta X_i')^{-1} X_i A_i \right\}^{-1} \sum_{i=1}^n A_i' X_i' (\sigma^2 I_{m_i} + X_i \Delta X_i')^{-1} Y_i \quad (3)$$

$$\underline{\beta}_1(a) = \begin{cases} [X_1' X_1 / \sigma^2 + \Delta^{-1} A_1' (\sum_{j=1}^n A_j' \Delta^{-1} A_j)^{-1} A_1' \Delta^{-1}]^{-1} X_1' Y_1 / \sigma^2 & \text{if } s \\ A_1 \underline{\beta}(a) & \text{if } s \end{cases} \quad (4)$$

where s denotes the set of sampled groups.

Note that $X_1' X_1$ does not have to be invertible for this estimation.

Pfeffermann (1978) shows the optimality of these estimators.

However, a different approach may be taken. The above approach minimizes the squared error loss over the sample with weights proportional to the reciprocal of the standard deviations. Instead, we can minimize the expected squared error loss in the prediction of Y values for the whole population. Furthermore, we can allow different weights for the different units in the population according to the sampling scheme or any other criterion. Therefore, we look for a statistic which is a linear function of the observed values $\underline{Y}_s$ and is unbiased, and will minimize the expected loss, i.e. a statistic $\underline{BY}_s$ such that

$$R(B) = \min_B E_s \sum_{i=1}^N \sum_{j=1}^{M_i} w_{ij} (Y_{ij} - X_{ij} \underline{BY}_s)^2 \quad (5)$$

where X_{ij} and w_{ij} are the known values of the explanatory variables and of the fixed weights respectively for the (i,j) - th unit. The expectation E_s is taken over the joint distribution of Y and the regression coefficients. The advantage of this loss function is that it minimizes the expected quadratic loss over the finite population and not only over the sample. After the overall estimate $\underline{BY}_s$ is obtained, separate estimates for the regression coefficients for each group can also be obtained. Their advantage is that each one combines information from all groups for which information is available in order to estimate the vector of coefficients of any one group

We now minimize $R(B)$. Note that:

$$\begin{aligned}
 R(B) &= E_{\xi} \sum_{i=1}^N \sum_{j=1}^{M_i} w_{ij} (Y_{ij} - X_{ij} \underline{B} \underline{Y}_{-s})^2 = \\
 &= E_{\xi} \sum_{i=1}^N \sum_{j=1}^{M_i} w_{ij} (Y_{ij} - X_{ij} \underline{\beta}_i)^2 + E_{\xi} \sum_{i=1}^N \sum_{j=1}^{M_i} w_{ij} (X_{ij} \underline{\beta}_i - X_{ij} \underline{B} \underline{Y}_{-s})^2 \\
 &= \sum_{i=1}^N \sum_{j=1}^{M_i} w_{ij} \sigma_i^2 + E_{\xi} R^*(B), \text{ (say),}
 \end{aligned} \tag{6}$$

since, assuming independence of $\underline{Y}_{-s}$ from Y_{ij} , conditional on $\underline{\beta}_i$

$$E_{\xi} \sum_{i=1}^N \sum_{j=1}^{M_i} w_{ij} (Y_{ij} - X_{ij} \underline{\beta}_i)(X_{ij} \underline{\beta}_i - X_{ij} \underline{B} \underline{Y}_{-s}) = 0.$$

Denoting the weighted design matrix of group i by

$$X_{wi} = (w_{i1}^{1/2} X_{i1}, \dots, w_{iM_i}^{1/2} X_{iM_i})' \text{ and } X_w' X_w = \text{diag}(X_{w1}' X_{w1}, \dots, X_{wN}' X_{wN})$$

we can rewrite

$$R^*(B) = [\underline{\beta}'_1 - \underline{Y}'_s B', \dots, \underline{\beta}'_N - \underline{Y}'_s B'] X_w' X_w [\underline{\beta}'_1 - \underline{Y}'_s B', \dots, \underline{\beta}'_N - \underline{Y}'_s B']'.$$

$$\text{Let } C(\tilde{\underline{\beta}}) = \left(\sum_{i=1}^N X_{wi}' X_{wi} \right)^{-1} \sum_{i=1}^N X_{wi}' X_{wi} \underline{\beta}_i.$$

Then

$$R^*(B) = [\underline{\beta}'_1 - C'(\tilde{\underline{\beta}}), \dots, \underline{\beta}'_N - C'(\tilde{\underline{\beta}})] X_w' X_w [\underline{\beta}'_1 - C'(\tilde{\underline{\beta}}), \dots, \underline{\beta}'_N - C'(\tilde{\underline{\beta}})]' \tag{7}$$

$$+ [(C'(\tilde{\underline{\beta}}) - \underline{Y}'_s B', \dots, (C'(\tilde{\underline{\beta}}) - \underline{Y}'_s B')] X_w' X_w [C'(\tilde{\underline{\beta}}) - \underline{Y}'_s B', \dots, C'(\tilde{\underline{\beta}}) - \underline{Y}'_s B']'$$

since the "middle" term is 0 by choice of $C(\tilde{\beta})$. Therefore, it is left to minimize

$$\begin{aligned} E_{\xi} R^{**}(B) &= E_{\xi} (C'(\tilde{\beta}) - \frac{Y' B'}{S}) (\sum_{i=1}^N X'_{w_i} X_{w_i}) (C(\tilde{\beta}) - BY_{\underline{S}}) \\ &= E_{\xi} \text{tr} [(C'(\tilde{\beta}) - \frac{Y' B'}{S}) (\sum_{i=1}^N X'_{w_i} X_{w_i}) (C(\tilde{\beta}) - BY_{\underline{S}})]. \end{aligned} \quad (8)$$

Since $BY_{\underline{S}}$ is required to be unbiased it follows that

$$E_{\xi} R^{*}(B) = \text{tr} [(\sum_{i=1}^N X'_{w_i} X_{w_i}) \text{Var} (BY_{\underline{S}} - C(\tilde{\beta}))].$$

The matrix $\sum_{i=1}^N X'_{w_i} X_{w_i}$ is non-negative definite, therefore its trace is non-negative. Therefore, we only need to find the unbiased estimator which brings $\text{Var}_{\xi}(BY_{\underline{S}} - C(\tilde{\beta}))$ to a minimum. This can be done using the following theorem from Pfeffermann (1978).

THEOREM: Let $L\tilde{\beta}$ be a set of linear combinations of $\tilde{\beta}$ where L is a known matrix. Let $BY_{\underline{S}}$ be a ξ - unbiased linear estimator of $L(\tilde{\beta})$ where the rows of B are a function of the appropriate rows in L . Then $\text{Var}_{\xi}(BY_{\underline{S}} - L(\tilde{\beta})) = \text{Var}_{\xi}(L(\tilde{\beta}(a)) - L\tilde{\beta}) + D$ where D is non-negative definite and $\tilde{\beta}(a)$ is given in (3) and (4).

According to the theorem, the optimal estimator is

$$B^*Y_{\underline{S}} = (\sum_{i=1}^N X'_{w_i} X_{w_i})^{-1} \sum_{i=1}^N X'_{w_i} X_{w_i} \tilde{\beta}_{-1}(a). \quad (9)$$

3. A Special Case: $A_i = I$

One of the more interesting cases is when we suspect that there are differences between the groups but we are completely ignorant of the patterns in these differences. In this case we assume all the β_i 's have exactly the same expected value or in other words $A_i \equiv I$ for $i = 1, \dots, N$. Then $E(\beta_i) = \beta$ and

$$\beta(a) = \left\{ \sum_{i=1}^n X_i' (\sigma^2 I_{m_i} + X_i \Delta X_i')^{-1} X_i \right\}^{-1} \sum_{i=1}^n X_i' (\sigma^2 I_{m_i} + X_i \Delta X_i')^{-1} Y_i \quad (10)$$

$$\beta_i(a) = \Delta X_i' (X_i \Delta X_i' + \sigma^2 I_{m_i})^{-1} Y_i + \{I - \Delta X_i' (X_i \Delta X_i' + \sigma^2 I_{m_i})^{-1} X_i\} \beta(a), \quad (11)$$

for sampled groups and $\beta_i(a) = \beta(a)$ for non-sampled groups.

One can also minimize the expected risk by use of Lagrange multipliers. Denote $X_s = \text{diag}(X_1, \dots, X_n)$ where X_i is the design matrix of the sampled units in the i -th group and X is the $M \times N(k+1)$ block diagonal matrix the i -th block of which is the design matrix of all the units in the i -th group, as denoted earlier. Then, choosing $w_{ij} = 1$ for all i, j and denoting $P_i = X_i' \Delta X_i + \sigma^2$

$$P_s = \text{diag}(P_1, \dots, P_n)$$

then after some algebraic manipulations we get

$$B^* \underline{Y}_s = (X'X)^{-1} \{ [X'X - X_s'X_s] [X_s'P_s^{-1} X_s]^{-1} X_s'P_s^{-1} + [X_s'X_s] [X_s'X_s]^{-1} X_s' \} \underline{Y}_s \quad (12)$$

This alternative version of (9) expresses our estimator as a weighted average of ordinary least squares estimators and generalized least squares estimators for the group.

It is usually the case that the design matrix is known only for the sample. If all groups are sampled, as for instance in a stratified sample, we can assume all units within each group have an equal probability of being selected. Denoting the selected ones as the first m_i units, we can in this case estimate the design matrix

$$X'_{w_i} X_{w_i} = \sum_{j=1}^{M_i} w_{ij} X'_{ij} X_{ij} \quad \text{by} \quad \frac{M_i}{m_i} \sum_{j=1}^{m_i} w_{ij} X'_{ij} X_{ij}$$

and substitute this into the estimator $\underline{B} \underline{Y}_{-s}$.

Another interesting aspect of the case $A_i = I$ is an examination of the limiting values. First note that

$$(X'_i \Delta X_i + \sigma^2 I_{m_i})^{-1} = 1/\sigma^2 I_{m_i} - X_i/\sigma^2 [X'_i X_i/\sigma^2 + \Delta^{-1}]^{-1} X'_i/\sigma^2 \quad (14)$$

and denote

$$C_i = X'_i X_i/\sigma^2 + \Delta^{-1} \quad (15)$$

Then, if $X'_i X_i$ are invertible for $i=1, \dots, n$, $B^* \underline{Y}_{-s}$ can be expressed as

$$\begin{aligned} B^* \underline{Y}_{-s} &= \left(\sum_{i=1}^N X'_{w_i} X_{w_i} \right)^{-1} \sum_{i=1}^N X'_{w_i} X_{w_i} [(\Delta X'_i X_i/\sigma^2 - (\Delta X'_i X_i/\sigma^2) C_i^{-1} X'_i X_i/\sigma^2) \hat{\beta}_i \\ &+ (I - \Delta X'_i X_i/\sigma^2 + (\Delta X'_i X_i/\sigma^2) C_i^{-1} X'_i X_i/\sigma^2) \underline{\beta}(a)] \end{aligned} \quad (16)$$

i.e. a weighted average of ordinary least squares estimators from the different groups and a generalized least squares estimator from the whole sample. Denoting $\Delta = \delta^2 \Delta_0$, where Δ_0 is fixed, we examine the cases $\delta^2/\sigma^2 \rightarrow 0$, and $\sigma^2/\delta^2 \rightarrow 0$, using the presentation (16).

We first examine the case $\delta^2/\sigma^2 \rightarrow 0$. When this occurs, estimation of $\underline{\beta}$ from the $\underline{\beta}_i$'s is fairly accurate but there are big errors in predicting Y using X. Therefore, intuitively, a small weight should be put on the separate ordinary least squares and a large weight should be put on the generalized least squares estimator. Inspecting the matrix of coefficients of $\hat{\underline{\beta}}_i$ for any $i=1, \dots, n$ we find that

$$\begin{aligned} \lim_{\delta^2/\sigma^2 \rightarrow 0} \Delta_0 X_i' X_i \delta^2/\sigma^2 - (\Delta_0 X_i' X_i \delta^2/\sigma^2) (X_i' X_i/\sigma^2 + 1/\delta^2 \Delta_0^{-1})^{-1} X_i' X_i/\sigma^2 \\ = \lim_{\delta^2/\sigma^2 \rightarrow 0} \Delta_0 X_i' X_i \delta^2/\sigma^2 - (\Delta_0 X_i' X_i \delta^2/\sigma^2) (X_i' X_i \delta^2/\sigma^2 + \Delta_0^{-1})^{-1} X_i' X_i \delta^2/\sigma^2 = 0. \end{aligned} \quad (17)$$

When $\sigma^2/\delta^2 \rightarrow 0$ the error of predicting Y using X is small compared to the error of the estimation of $\underline{\beta}$ from the $\underline{\beta}_i$'s. Therefore, increasing weight should be put on the ordinary least squares estimators. Examining again the matrix of coefficients of $\hat{\underline{\beta}}_i$, and using (14) and (15) to substitute C_i^{-1} we find

$$\begin{aligned} \lim_{\sigma^2/\delta^2 \rightarrow 0} \Delta X_i' X_i/\sigma^2 - \Delta X_i' (1/\sigma^2 I_{m_i} - (X_i \Delta X_i' + \sigma^2 I_{m_i})^{-1}) X_i \\ = \lim_{\sigma^2/\delta^2 \rightarrow 0} \delta^2 \Delta_0 X_i' [\delta^2 (X_i \Delta_0 X_i' + \sigma^2/\delta^2 I_{m_i})]^{-1} X_i = I. \end{aligned}$$

Another interesting case is when $\Delta=0$ or when the model is an ordinary linear regression model with uncorrelated, equal variance

errors, but we assume the model of full exchangeability between groups is true and estimate accordingly. Then the estimation of Δ will give a nearly-singular matrix, and a natural step is to substitute $\Delta=0$ in the model. This reduces the model to a linear regression model. It can be easily seen that substituting $\Delta=0$ in B^*Y_s given in (9), (10), (11) or (16) gives the ordinary least squares estimator $\hat{\beta}$ for all β_1 and for β . Similarly, if we assume a different variance for the errors in each group B^*Y_s becomes a generalized least squares estimator for β . This supports the generality and optimality of B^*Y_s since in simple models it reduces to well known optimal estimators.

A proposed estimator in stratified sampling is the Klein estimator. This estimator of β weights each observation on every variable by the squared root of the reciprocal of its probability of inclusion. If $w_{ij}=1$ for $i=1, \dots, N$, $j=1, \dots, M_i$, the estimator is

$$\underline{\beta}(K) = \left(\sum_{i=1}^N M_i / m_i \times X_i' X_i \right)^{-1} \sum_{i=1}^N M_i / m_i \times X_i' Y_i. \quad (19)$$

If $\delta^2/\sigma^2 \rightarrow 0$ in B^*Y_s and the weights are taken to be M_i/m_i , B^*Y_s simplifies to the Klein estimator. This shows that the Klein estimator is nearly optimal in the sense of minimizing the expected squared error loss over the population when the design matrix is only known for the sample.

One of the components in $\underline{\beta}(a)$ is

$$D = \left[\sum_{i=1}^n X_i' (X_i \Delta X_i' + \sigma^2 I_{m_i})^{-1} X_i \right]^{-1} \quad (20)$$

Assuming Δ is of full rank and that $(X_i \Delta X_i' + \sigma^2 I_{m_i})$ is invertible for all $i=1, \dots, n$ it is clear that the sum in D has to span a k-dimensional vector space for D to be well-defined, but not every summand has to be k-dimensional. By ordering the observations in each group appropriately, the design matrices of the different groups would be presented so that their sum spans $\bigcup_{i=1}^n M(X_i)$ where $M(X_i)$ is the vector space spanned by X_i . This defines the conditions for D to be well defined.

In the most extreme case, each observation is a group. Then for D to be well defined, it is necessary that there be at least k observations and that at least k of the observed vectors be linearly independent of each other. Given this condition estimation as presented in (9), (10) and (11), is possible. Ordinary least squares is not possible here, because it requires that $\text{rank}(X_i) \geq k$ for all $i=1, \dots, n$.

Another approach which may be taken to the problem of finding a single estimator is to minimize the risk only over the units of the population which were not sampled. This approach supposes that there is no need to predict Y for those units already sampled. In this case the weights w_{ij} for the sampled units are 0. Estimation proceeds as above with (10), (11) remaining unchanged and (9) is changed according to the new weights.

4. Variance Estimation

The estimators presented so far assumed that the quantities σ^2 and Δ are known (see e.g. (3), (4), (10), (11)). These quantities are usually unknown and must be estimated from the sample. The variance σ^2 is the usual variance of the random variable Y , given the values of the explanatory variables and their corresponding regression coefficients, β_{-1} , i.e. disregarding the randomness of those coefficients. The variance matrix Δ reflects the variance of the regression coefficients around their expected values. Two methods for estimating σ^2 and Δ , applicable to the models presented can be considered.

The first method estimates the variance components one by one. This is an extension of a variance estimation method presented in Pfeffermann (1978) for the case $A_i = I$ and Pfeffermann and Nathan (1981) for the univariate case. First, σ^2 is estimated via an ordinary least squares estimation in each group. If the matrix $X_i'X_i$ is not invertible for some groups we can take a generalized inverse matrix. The residual sum of squares is the same, whatever generalized inverse is chosen.

Then

$$\hat{\sigma}^2 = \frac{\text{RSS}}{\sum_{i=1}^n m_i - n(k+1)} = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (Y_{ij} - \hat{Y}_{ij})^2}{\sum_{i=1}^n m_i - n(k+1)} \quad (21)$$

where k is the number of explanatory variables, as before.

Next, we note that the ordinary least squares estimator of $\underline{\beta}_i$ is $\underline{\beta}_{-1} = (X_i'X_i)^{-1}X_i'Y_i$ (when $(X_i'X_i)^{-1}$ exists)--- has variance:

$$\text{Var}(\hat{\underline{\beta}}_i) = \Delta + \sigma^2(X_i'X_i)^{-1} \quad i=1, \dots, n \quad (22)$$

Denoting the $(1,1^1)$ element of Δ by $\delta_{1,1^1}$ ($1,1^1 = 0,1,\dots,k$) and $\delta_{1,1^1} = \delta_1^2$ we proceed to estimate δ_0^2 . We consider the model

$$\hat{\beta}_{i0} = A_{i0}\underline{\beta} + \eta_i \quad (i=1,\dots,n) \quad (23)$$

where A_{i0} denotes the first row of A_i . Then

$$\text{Var}(\eta_i) = \text{Var}(\hat{\beta}_{i0}) = \delta_0^2 + \sigma^2(X_i'X_i)^{-1}_{0,0}$$

$$\text{Cov}(\eta_i, \eta_j) = \text{Cov}(\hat{\beta}_{i0}, \hat{\beta}_{j0}) = 0 \quad i \neq j.$$

and $\underline{\beta}$ can be estimated from (23) using generalized least squares.

Denoting

$$\lambda_{i0} = \delta_0^2 [\delta_0^2 + \sigma^2(X_i'X_i)^{-1}_{0,0}]^{-1} \quad (24)$$

we obtain as the estimator of $\underline{\beta}$ on the basis of $\{\hat{\beta}_{i0}\}$: (25)

$$\tilde{\underline{\beta}}_0 = \begin{bmatrix} \sum_{i=1}^n \lambda_{i0} a_{i1}^2 & \sum_{i=1}^n \lambda_{i0} a_{i1} a_{i2} & \dots & \sum_{i=1}^n \lambda_{i0} a_{i1} a_{ir} \\ \cdot & \cdot & \cdot & \cdot \\ \sum_{i=1}^n \lambda_{i0} a_{i1} a_{is} & & \sum_{i=1}^n \lambda_{i0} a_{ir}^2 & \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n \lambda_{i0} \hat{\beta}_{i0} a_{i1} \\ \cdot \\ \sum_{i=1}^n \lambda_{i0} \hat{\beta}_{i0} a_{ir} \end{bmatrix}$$

Now we examine the model

$$\hat{\beta}_{i0} \sqrt{\lambda_{i0}} = A_{i0} \sqrt{\lambda_{i0}} \underline{\beta} + \lambda_{i0} \eta_i \quad i=1,\dots,n. \quad (26)$$

This model is homoscedastic with the variance of $\sqrt{\lambda_{10}}\eta_i$ exactly equal to δ_0^2 . Then the estimate of δ_0^2 is obtained as the non-negative solution of the equation

$$\delta_0^2 = \frac{RSS}{n-1} = \frac{1}{n-1} \sum_{i=1}^n \hat{\lambda}_{10} (\hat{\beta}_i - A_{10} \tilde{\beta}_0)^2 \quad (27)$$

where $\hat{\lambda}_{10}$ is obtained by substituting $\hat{\delta}_0^2$ for δ_0^2 and $\hat{\sigma}^2$ for σ^2 in (24), and replaces λ_{10} in (25) to obtain $\tilde{\beta}_0$. Equation (27) is solved iteratively and an estimate $\hat{\delta}_0^2$ is thus obtained. The process is repeated for $\delta_1^2, \dots, \delta_k^2$ and the estimates $\hat{\delta}_1^2, \dots, \hat{\delta}_k^2$ are thus obtained. Note that the estimate $\tilde{\beta}_0$ in (25) is different for each of the explanatory variables involved.

It is left to estimate δ_{ij} for $i \neq j$.

Note that

$$Z_{1l} = \frac{\hat{\beta}_{1l} - A_{1l} \tilde{\beta}_1}{\hat{\sigma}(\hat{\beta}_{1l})} \quad \begin{array}{l} i=1, \dots, n \\ l=0, \dots, k \end{array}$$

are all random variables which have mean 0 and variance 1. Examining the correlation between any two of these variables that belong to the same group, we find that

$$\text{Cov}(Z_{1l}, Z_{1m}) = \frac{\text{Cov}(\hat{\beta}_{1l}, \hat{\beta}_{1m})}{\hat{\sigma}(\hat{\beta}_{1l}) \hat{\sigma}(\hat{\beta}_{1m})} + \frac{\delta_{1,m}}{\hat{\sigma}(\hat{\beta}_{1l}) \hat{\sigma}(\hat{\beta}_{1m})} \quad \begin{array}{l} i=1, \dots, n \\ l \neq m \end{array} \quad (28)$$

The covariance in (28) is $\sigma^2 (X_i' X_i)^{-1}_{1,m}$ and $\delta_{1,m}$ can be estimated by equating the averages of the observed and of the expected value of Z_{i1}, Z_{im} over all the groups, the estimators being the solutions of the equations

$$\sum_{i=1}^n \frac{(\hat{\beta}_{i1} - A_{i1} \tilde{\beta}_1)(\hat{\beta}_{im} - A_{im} \tilde{\beta}_m)}{\hat{\sigma}(\hat{\beta}_{i1}) \hat{\sigma}(\hat{\beta}_{im})} = \sum_{i=1}^n \frac{\sigma^2 (X_i' X_i)^{-1}_{1,m}}{\hat{\sigma}(\hat{\beta}_{i1}) \hat{\sigma}(\hat{\beta}_{im})} + \frac{\delta_{1,m}}{\hat{\sigma}(\hat{\beta}_{i1}) \hat{\sigma}(\hat{\beta}_{im})} \quad (29)$$

for $0 \leq 1 < m \leq k$.

In the case $A_i = I$, the estimators $\hat{\delta}_0^2, \dots, \hat{\delta}_k^2$ are consistent and asymptotically normal (Pfeffermann (1978)).

Also, if we use these estimators to test hypotheses of the type $H_0: \delta_1^2 = 0$, even though the probability of type I error is high, the estimate of δ_1^2 will be very small. These properties may extend to the general case ($A_i \neq I$) but this has not been shown.

The advantage of this method of estimation is its simplicity. It can be performed easily by inverting at most $(k+1) \times (k+1)$ or $r \times r$ matrices. All the variances involved have non-negative estimates by construction. However, the method has some drawbacks. One is that σ^2 is never updated after the initial estimation, which does not account for the full complexity of the model. Secondly, we first estimate variances from the observations independently, thus tacitly assuming that the covariances are all zero, and then proceed to estimate those same covariances. One way to overcome both these problems is to exploit the EM algorithm - Dempster, Laird and Rubin (1978). The basic idea of the EM algorithm is to view the actual data, in our case $\underline{Y}_i, i=1, \dots, n$, as incomplete data from a larger data set, in our case

$\underline{Y}_i, \underline{\beta}_i \quad i=1, \dots, n$. An EM algorithm is an iterative procedure, each iteration of which consists of the two steps, called the E-step and the M-step. The M-step simulates the maximum likelihood estimation which would be carried out if the data were complete, but since the data are incomplete the maximization of the M-step can only be carried out on a proxy for the likelihood, where the proxy is produced by the E-step. To be precise the E-step finds the posterior expectation of the sufficient statistics based on the complete data, given the actual data $(\underline{Y}_i, i=1, \dots, n)$ and given the current estimates of the parameters $(\sigma^2 \text{ and } \Delta)$. The M-step then uses the expected sufficient statistics to produce the next estimated values of σ^2 and Δ .

The following derivation by the EM method for estimating σ^2 and Δ is based on Dempster, Rubin and Tsutakawa (1981). The basic approach assumes the model is a mixed model of fixed and random effects and applies Bayesian principles in the analysis. Rewriting the basic model as

$$\underline{\beta}_i = A_i \underline{\beta} + \underline{\eta}_i \quad (30)$$

where $E(\underline{\eta}_i) = 0$ for $i=1, \dots, n$, the model can be written as

$$\underline{Y} = X^*(\underline{\beta}'; \underline{\eta}')' + \underline{\varepsilon} \quad (31)$$

where X^* is the appropriate design matrix and the errors $\underline{\varepsilon}$ are independently $N(0, \sigma^2)$ distributed. Assuming a flat prior for $\underline{\beta}$ (the MLR approach) and that a priori $\underline{\eta}_i \sim \eta(0, \Delta)$ independently of each other, we obtain a normal posterior distribution for the vector of coefficients. Denoting the residuals from estimation, $Y_{ij} - \hat{Y}_{ij}$, as e_{ij} for $i=1, \dots, n, j=1, \dots, m_i$ the complete data sufficient statistic in this case is $\sum_{i=1}^n \sum_{j=1}^{m_i} e_{ij}^2$

and the matrix $\underline{\eta} \underline{\eta}^1$. Starting with some initial value $(\Delta^{(0)}, \sigma^{(0)2})$ of (Δ, σ^2) , the algorithm essentially consists of iteratively estimating the sufficient statistic with an E-step and modifying the value of (Δ, σ^2) with an M step until convergence is attained.

In the E-step, given $(\Delta^{(i)}, \sigma^{(i)2})$ we first compute the posterior mean and covariance matrix of $(\underline{\beta}^1, \underline{\eta}^1)$ and denote

$$\begin{aligned} \underline{Y}^{(i)} &= (\underline{\beta}^{(i)1}, \underline{\eta}^{(i)1})^1 = P^{(i)-1} X^{*1} Y \\ \underline{\Sigma}_{\underline{Y}}^{(i)} &= \sigma^{(i)2} P^{(i)-1} \end{aligned} \quad (32)$$

where

$$P^{(i)} = X^{*1} X + \sigma^{(i)2} \text{diag}(0; \Delta^{(i)-1}, \dots, \Delta^{(i)-1})$$

and $P^{(i)-1}$ exists and is defined by

$$P^{(i)-1} = \begin{bmatrix} H^{-1} & G_1 & \dots & G_n \\ G_1^1 & D_{11} & D_{12} & \dots & D_{1n} \\ \vdots & & D_{12} & D_{22} & \\ \vdots & & & \ddots & \\ G_n^1 & D_{1n} & & & D_{nn} \end{bmatrix} \quad (33)$$

We then estimate the sufficient statistic by the conditional expectations.

$$E(\underline{\eta} \underline{\eta}^1 \mid \underline{Y}, \Delta^{(i)}, \sigma^{(i)2}) = \underline{\eta}^{(i)} \underline{\eta}^{(i)1} + \underline{\Sigma}_{\underline{\eta}}^{(i)} \quad (34)$$

and

$$E(\underline{e}'\underline{e} \mid \underline{y}, \Delta^{(i)}, \sigma^{(i)^2}) = \underline{y}'\underline{y} - 2\underline{y}'\underline{X}^* \underline{\gamma}^{(i)} + \underline{\gamma}^{(i)'} \underline{X}^{*'} \underline{X}^* \underline{\gamma}^{(i)} + \text{tr}(\underline{X}^{*'} \underline{X}^* \underline{\Sigma}_{\underline{Y}}^{(i)}). \quad (35)$$

In the M-step, we simulate maximum likelihood by using the conditional expectations (34) and (35), in lieu of $\underline{\eta}\eta'$, and $\underline{e}'\underline{e}$ to obtain $(\Delta^{(i+1)}, \sigma^{(i+1)^2})$,

$$\sigma^{(i+1)^2} = \{ \underline{y}'\underline{y} - 2\underline{y}'\underline{X}^* \underline{\gamma}^{(i)} + \underline{\gamma}^{(i)'} \underline{X}^{*'} \underline{X}^* \underline{\gamma}^{(i)} + \text{tr}(\underline{X}^{*'} \underline{X}^* \underline{\Sigma}_{\underline{Y}}^{(i)}) \} / n \quad (36)$$

$$\Delta^{(i+1)} = \frac{\sum_{j=1}^n \underline{\eta}_j^{(i)} \underline{\eta}_j^{(i)'}}{n} + \frac{\sum_{j=1}^n D_{jj}^{(i)} \sigma^{(i)^2}}{n}.$$

The advantage of the EM algorithm is that it estimates and updates jointly σ^2 and Δ and it has the appealing feature of being based on maximum likelihood, thus the likelihood is increased at each step. Its drawback is that computationally it is more difficult although, according to Dempster *et al.* (1981) a timing comparison shows that EM can be quite competitive with the Newton-Raphson method. In our data set, the first method of estimation proposed converged in much fewer iterations than the EM algorithm method.

5. The Analysis of the Data

The theory developed in this project was applied to data from the Citrus Branch Profitability Survey 1979, carried out by the Institute of Farm Income Research. The objective of the analysis is to explain productivity and profitability in this branch using both "physical" explanatory variables such as amount of water, fertilizers and "economic" variables such as direct input, work days in cultivating. The analysis pertains only to the grapefruit crop and not to the citrus branch in general. Even though other crops in the citrus branch may be similar, it is probable that there are also inherent differences. Therefore, a separate analysis for each crop or an extensive multivariate treatment might be more efficient.

The data is naturally divided into groups (strata) according to geographical regions and type of settlement. Preliminary analysis screened out some variables. Several variables were connected with productivity in the sense that productivity "explained" them instead of vice versa. For example, the number of workdays in picking is hardly a criterion for explaining productivity, but rather a result of productivity. It was also clear from the preliminary analysis that the size of the farm influenced the values of all variables very strongly. Therefore, it was decided that wherever relevant, the value of the variable would be divided by the area. Thus, we obtain, for example, production per dunam as explained by water per dunam.

The variables measured by the Institute of Farm Income Research and which are relevant for this analysis are as follows (several new variables were constructed on the basis of the data and are also given):

1. Area of grapefruit orchard.
2. Productivity (amount of fruit) per dunam.
3. Productivity value per dunam.
4. Direct input (in thousands of IL 1979) per dunam.
5. Administrative costs per dunam.
6. Depreciation costs per dunam.
7. Expenses for hired labor per dunam.
8. Expenses for owner's labor per dunam.
9. Interest payments per dunam.
10. Profit/loss per dunam.
11. Expenses on equipment per dunam.
12. Expenses on organic fertilizers per dunam.
13. Expenses on chemical fertilizers per dunam.
14. Expenses on microelements fertilizers per dunam.
15. Expenses on pesticides per dunam.
16. Other expenses per dunam.
17. Interest on operating capital per dunam.
18. Amount of pure nitrogen per dunam.
19. Amount of organic fertilizers per dunam.
20. Amount of water per dunam.
21. Amount of purchased water per dunam.
22. Revaluation capital per dunam.
23. Revaluation of orchard per dunam.
24. Expenses on hired work in cultivating per dunam.
25. Expenses on management and supervision per dunam.
26. Age of orchard.
27. Sample inflation factor.

The distribution of the observations according to the groups is given in Table 1.

Type of farm: <u>Region</u>	Qibbuz or <u>Moshav</u>	<u>Table 1</u> Commercial cultivation <u>companies</u>				<u>Cooperative</u>	<u>Private</u>	<u>Total</u>
		<u>Family</u>						
Inner valleys	28	2	2		7	-		39
Northern coastal plain	11	3	-		2	-		16
Yizre'el valley	11	5	-		3	-		19
Binyamina-Tel Aviv	8	13	11		7	6		45
Rehovot-Ashkelon	10	13	17		15	3		58
Lakhish and Negev	6	-	-		10	-		16
Other	4	-	1		5	-		10
Total	78	36	31		49	9		203

Two observations were clear outliers and are not included in Table 1.

Note that in Table 1 there are 10 observations which do not belong to any of the regions (denoted "other"). These observations were excluded from the analysis since they were not assumed to form a homogeneous group and in any case there were too few observations in these groups. The distributions of all the variables and their intercorrelations were examined. Many of the variables are distributed asymmetrically with much more mass at the lower values. Special emphasis was put on the examination of the sample inflation factor. It was noticed that the inflation factor is not strongly correlated with any of the explanatory variables. Other variables, though, were sometimes strongly correlated with each other. For ease of interpretation, the procedure was to drop

one of a pair of strongly correlated variables. For example, interest payments on capital have a strong negative correlation with age and therefore, the variable of interest payments on capital was dropped. The variables (and their abbreviations which will be used throughout) that ultimately remained in the analysis are:

1. Direct inputs (DI).
2. Administrative costs (AC).
3. Depreciation (DP).
4. Expenses on hired labor (not including picking)(HW).
5. Expenses on owner's labor (not including picking) (SW).
6. Age of orchard (AGE).
7. Expenses on mechanical equipment (ME).
8. Expenses on pesticides (PE).
9. Amount of pure nitrogen (PN).
10. Expenses on microelements (MCE).
11. Other expenses (OE).
12. Amount of organic fertilizers (OF).
13. Amount of water (AW).
14. Sample inflation factor (INF).

At the first stage an overall linear regression model was tried on the two dependent variables-production per dunam and production value per dunam. The result was $R^2 = .12501$, $R^2_{adj} = .05620$ for the former, with the coefficients of age and direct input significant, and $R^2 = .28982$, $R^2_{adj} = .2339$ for the latter, with the coefficients of depreciation, age, direct inputs, amount of water and expenses on hired labor (negative) being significant. Since this didn't seem good enough (too small a part

of the variability of the dependent variable explained by the explanatory variables), it was decided to separate the observations into groups according to region and type of farm. Then, it was proposed to fit different models to the different groups with the intent of combining information from the different groups. A stepwise regression was carried out for each group with sufficient observations. The results showed that different variables entered the regressions in the different groups. Regressions were also performed on a "production function" type model, i.e.

$$Y = X_1^{\beta_1} \dots X_k^{\beta_k} \epsilon.$$

A problem, here was that for some variables there are many zero observations. In the most extreme case, the amount of microelements was zero for 60% of the observations. It was decided to dichotomize this variable, thus making it more easily interpretable. In order to carry out the analysis every zero was changed to be 10^{-6} . The smallest order of magnitude for non-zero observations was 10^{-3} . (The multiplicative model was temporarily abandoned both because of technical difficulties of too many zeros and because it did not appear at this point that a greater part of the variability in the production was explained by it.) The models obtained using stepwise regression involved overall all 15 variables, but for each group there were only about 4-6 variables in the regression equation. This is partially due to the small sample sizes within groups (10-20 observations usually).

In order to decide which are the more important variables over all groups, it was decided to take different sets of variables which appeared in one or more regressions and enter each set of them into regressions for all the groups and then average the R^2 . Three types of averaging were used. The first is a simple average of the correlations

from all the groups, the second and third are weighted averages of the correlations with firstly the sample sizes of the groups as weights, and secondly the inflating factors as weights. The multiple correlations from the regressions using a linear model are shown in Tables 2 (a) & (b). The results were similar for the multiplicative model. The number of variables in a set, (4 for production per dunam and 3 for production value per dunam) was chosen on the basis of the number of significant coefficients in the previous stepwise regression.

As can be seen, the "optimal" set of variables for explaining production per dunam is the one which includes area; expenses on mechanical equipment; direct input and expenses on pesticides. The "optimal" set of variables for explaining production value per dunam is the one which includes age; depreciation costs and administrative expenses. Next a pattern in the coefficients of the variables across the groups was sought. These coefficients are presented in Table 3(a) when the dependent variable is production per dunam and in Table 3(b) when the dependent variable is production value per dunam. Most of these coefficients were not significant, since there were too few observations in each group. Upon inspection, though, it appears that the coefficients of the Binyamina-Tel Aviv region are different from those of the Rehovot-Ashqelon region, and observations from those two regions should not be combined. The same is true for the Northern Coastal Plain and the Yizre'el Valley.

Table 2(b)

R² For separate regressions on different groups and set of variables

Dependent variable - Production value per dunam

Variables in Regression		Sum of sample inflation factor	Sample Size	Type of Farm	R ²																
Region					AC DP	AC HW	AC Age	AC SW	AC Age	AC HW	AC Age	AC SW	AC ME	AC DP	AC PN	AC ME	AC DP	AC PN	Area PE	Area D1	Area MCE
Inner valleys	Qibbuz or Moshav	28	47.38		.13	.16	.19	.25	.21	.22	.13	.13	.13	.03	.55	.09	.04	.06			
Northern coastal plain	Qibbuz or Moshav	11	16.19		.57	.42	.26	.29	.39	.37	.37	.29	.15	.36	.37	.32	.35	.32			
Yizre'el valley	Qibbuz or Moshav	11	18.65		.33	.46	.35	.49	.32	.35	.32	.46	.36	.42	.48	.17	.46	.13			
Binyamina-Tel Aviv	Qibbuz or Moshav	8	46.90		.66	.09	.22	.22	.14	.44	.17	.21	.59	.63	.09	.97	.82	.69			
Rehovot-Ashqelon	Qibbuz or Moshav	10	26.16		.68	.73	.75	.75	.47	.74	.69	.97	.87	.55	.51	.44	.10	.68			
Northern coastal plain	Family	8	211.72		.42	.34	.45	.41	.35	.55	.43	.41	.38	.35	.83	.27	.50	.21			
Binyamina-Tel Aviv	Family	13	1642.81		.09	.07	.10	.08	.11	.20	.11	.09	.11	.11	.35	.50	.35	.04			
Rehovot-Ashqelon	Family	13	923.92		.39	.51	.37	.29	.35	.30	.39	.31	.38	.06	.10	.12	.04	.17			
Inner valleys	Commercial Cultivation companies	9	10.84		.94	.13	.41	.31	.39	.79	.78	.78	.70	.76	.15	.46	.77	.13			
Binyamina-Tel Aviv	"	11	596.04		.27	.23	.39	.39	.12	.64	.21	.34	.30	.27	.40	.23	.37	.27			
Rehovot-Ashqelon	"	17	739.14		.17	.19	.19	.08	.13	.09	.21	.13	.22	.20	.10	.24	.34	.17			
Binyamina-Tel Aviv	Cooperative	7	99.09		.71	.48	.74	.66	.73	.80	.51	.63	.85	.34	.45	.66	.51	.71			
Rehovot-Ashqelon	Cooperative	15	136.48		.04	.05	.18	.15	.17	.19	.17	.02	.17	.14	.18	.37	.07	.21			
Lakhish and Negev	Cooperative	10	47.56		.38	.20	.21	.21	.14	.25	.26	.27	.31	.14	.27	.25	.23	.17			
Binyamina-Tel Aviv	Private	9	758.62		.41	.46	.51	.53	.34	.51	.39	.39	.38	.17	.04	.13	.32	.13			
Average 1	$\Sigma R_1^2 / 15$				.41	.30	.35	.34	.29	.43	.34	.36	.39	.30	.32	.35	.35	.27			
Average 2	$\Sigma m_1 R_1^2 / \Sigma m_1$				.35	.28	.32	.31	.26	.37	.30	.31	.33	.25	.33	.31	.30	.23			
Average 3	$\Sigma w_1 R_1^2 / \Sigma w_1$				.26	.26	.29	.25	.21	.33	.25	.24	.27	.17	.25	.30	.30	.15			

Table 3(a) Regression coefficients from o.l.s.

Coefficients of Area

<u>Region</u>	<u>Qibbuz or Moshav</u>	<u>Family</u>	<u>Commercial/ Cultivation Companies</u>	<u>Cooperative</u>	<u>Private</u>
Inner Valleys	-0.00146	0	-0.00013	-	-
Northern coastal plain	0.01142	-0.04184	-	0	-
Yizre'el Valley	-0.00211	-	-	0	-
Binyamina-Tel Aviv	-0.00809	0.00418	-0.00761	0.02243	0.06056
Rehovot-Ashqelon	-0.00352	-0.02323	-0.00302	0.00155	-
Lakhish and Negev	0.00636	-	-	0.0225	-

Coefficients of direct inputs excluding pesticides

Inner Valleys	-0.158	0	1.333	-	-
Northern coastal plain	0.36	3.075	-	0	-
Yizre'el Valley	0.199	-	-	0	-
Binyamina-Tel Aviv	0.015	2.083	2.559	2.386	0.943
Rehovot-Ashqelon	0.754	3.116	2.85	3.131	-
Lakhish and Negev	1.755	-	-	0.43	-

Coefficients of Pesticides

Inner Valleys	1.227	0	0.797	-	-
Northern coastal plain	1.237	-1.366	-	0	-
Yizre'el Valley	-0.882	-	-	0	-
Binyamina-Tel Aviv	9.718	6.019	4.214	6.091	2.543
Rehovot-Ashqelon	-1.654	-2.611	2.312	-6.155	-
Lakhish and Negev	1.11	-	-	-1.769	-

Coefficients of expenses on mechanical equipment

Inner Valleys	0.039	0	-2.778	-	-
Northern coastal plain	0.098	-8.899	-	0	-
Yizre'el Valley	0.590	-	-	0	-
Binyamina-Tel Aviv	-5.027	-4.008	-5.912	-13.337	8.366
Rehovot-Ashqelon	1.003	-9.078	-1.851	11.457	-
Lakhish and Negev	-5.851	-	-	13.43	-

Table 3(b) Regression coefficients from o.l.s.

<u>Coefficients for administrative expenses</u>					
<u>Region</u>	<u>Qibbuz or Moshav</u>	<u>Family</u>	<u>Commercial/ Cultivation Companies</u>	<u>Cooperative</u>	<u>Private</u>
Inner Valleys	6.245	0	-7.503	-	-
Northern coastal plain	4.518	15.293	-	0	-
Yizre'el Valley	4.577	-	-	0	-
Binyamina-Tel Aviv	9.882	4.429	-15.506	20.167	-
Rehovot-Ashqelon	14.122	18.549	-7.327	-0.833	-16.279
Lakhish and Negev	29.185	-	-	-9.934	-

<u>Coefficients for depreciation costs</u>					
Inner Valleys	1.993	0	12.659	-	-
Northern coastal plain	8.319	-13.135	-	0	-
Yizre'el Valley	2.824	-	-	0	-
Binyamina-Tel Aviv	-20.627	8.376	6.286	9.885	-3.069
Rehovot-Ashqelon	-1.862	1.888	1.776	3.692	-
Lakhish and Negev	-55.615	-	-	2.469	-

<u>Coefficients for age</u>					
Inner Valleys	-0.134	0	-0.037	-	-
Northern coastal plain	0.245	0.378	-	0	-
Yizre'el Valley	-0.105	-	-	0	-
Binyamina-Tel Aviv	0.491	0.195	0.258	0.491	0.081
Rehovot-Ashqelon	0.147	-0.076	-0.011	0.141	-
Lakhish and Negev	-0.165	-	-	-0.076	-

legend: - no observations in group.

0 too few observations in group for o.l.s.

Table 4(a) Regression coefficients using fixed effects

<u>Region</u>	<u>Coefficients of Area</u>				
	<u>Qibbuz or Moshav</u>	<u>Family</u>	<u>Commercial/ Cultivation Companies</u>	<u>Cooperative</u>	<u>Private</u>
Inner Valleys	-0.0025	0.0032	-0.0019	0.0034	-
Northern coastal plain	0.0357	0.0414	-	0.0416	-
Yizre'el Valley	-0.0048	0.0009	-	0.0011	-
Binyamina-Tel Aviv	0.0009	0.0066	0.0015	0.0068	0.0595
Rehovot-Ashqelon	-0.0010	0.0046	-0.0004	0.0049	0.0576
Lakhish and Negev	-0.0046	-	-	0.0013	-

<u>Coefficients of Expenses on total direct inputs excluding pesticides</u>					
Inner Valleys	0.436	3.081	4.896	0.514	-
Northern coastal plain	0.380	3.025	-	0.330	-
Yizre'el Valley	0.368	3.013	-	0.446	-
Binyamina-Tel Aviv	1.44	4.091	5.906	1.524	4.299
Rehovot-Ashqelon	1.212	3.857	5.672	1.290	4.065
Lakhish and Negev	2.172	-	-	2.250	-

<u>Coefficients of Expenses on mechanical equipment</u>					
Inner Valleys	-0.382	-1.265	-2.578	7.284	-
Northern Coastal plain	-1.924	-2.806	-	5.743	-
Yizre'el Valley	-2.619	-3.501	-	5.048	-
Binyamina-Tel Aviv	-7.265	-8.147	-9.46	0.402	-3.629
Rehovot-Ashqelon	0.365	-0.498	-1.811	8.051	4.02
Lakhish and Negev	-11.764	-	-	-4.097	-

<u>Coefficients of Expenses on pesticides</u>					
Inner Valleys	2.907	-0.467	-3.421	-2.096	-
Northern coastal plain	3.649	0.275	-	-1.354	-
Yizre'el Valley	2.371	-1.003	-	-2.632	-
Binyamina-Tel Aviv	10.441	7.067	4.113	5.438	-3.736
Rehovot-Ashqelon	-0.886	-4.260	-7.214	-5.889	-15.063
Lakhish and Negev	-1.415	-	-	-6.418	-

Table 4(b) Regression coefficients using fixed effects

Coefficients of Administrative expenses

<u>Region</u>	<u>Qibbuz or Moshav</u>	<u>Family</u>	<u>Commercial/ Cultivation Companies</u>	<u>Cooperative</u>	<u>Private</u>
Inner Valleys	0.956	-0.616	-5.849	-2.019	-
Northern coastal plain	4.241	2.669	-	1.266	-
Yizre'el Valley	1.244	-0.328	-	-1.731	-
Binyamina-Tel Aviv	3.283	1.711	-3.522	0.308	-1.179
Rehovot-Ashqelon	4.984	3.412	-1.821	2.009	0.522
Lakhish and Negev	1.484	-	-	-1.491	-

Coefficients of Depreciation costs

Inner Valleys	-0.359	0.663	1.714	1.834	-
Northern coastal plain	0.666	1.688	-	2.859	-
Yizre'el Valley	-0.374	0.648	-	1.819	-
Binyamina-Tel Aviv	1.558	2.565	3.616	3.736	-0.204
Rehovot-Ashqelon	-1.341	-0.319	0.732	0.852	-3.088
Lakhish and Negev	0.735	-	-	2.928	-

Coefficients of Age

Inner Valleys	-0.018	-0.098	-0.039	-0.008	-
Northern coastal plain	0.185	0.105	-	0.195	-
Yizre'el Valley	0.078	-0.002	-	0.088	-
Binyamina-Tel Aviv	0.145	0.065	0.124	0.155	0.017
Rehovot-Ashqelon	0.052	-0.028	0.031	0.062	-0.076
Lakhish and Negev	-0.034	-	-	-0.024	-

It should be emphasized that the results in Table 3 do not necessarily imply that the coefficients are different for the different regions or types of farm. The coefficients are not significant due to the size of the groups and may well be similar. Not combining the observations from different regions or types of farm is merely a precaution in case we find later that some coefficients of some of the variables are different for different regions or types of farm.

We were also concerned whether there is a systematic relation between any of the regression coefficients ($\hat{\beta}_i$) and some of the variable averages, $\bar{X}_{ij}$, across regions and types of farms. For example we checked whether the coefficient of direct input increases with age. We found no strong relations which required a transformation of any of the variables. The techniques employed were plots, examining R^2 and canonical correlation analysis.

The underlying assumption, as explained earlier, is that the same variables explain the dependent variable although the coefficients may be different in the different groups. The differences in Tables 3(a) and (b) could be due to the size of the groups, as noted. However, another possible explanation for these differences are that the effects are fixed but different for the different regions and different types of farms. This calls for a regression model whose coefficients can be additively decomposed into fixed effects for the different regions and fixed effects for the types of farms. Denoting by $Y_{ij k}$ the k-th unit in the group of farms of type j in region i, by μ "grand mean" effects, by α_i the region effects and by γ_j the type of farm effects, the model is

$$Y_{ij k} = (\mu + \alpha_i + \gamma_j)' X_{ij k} + e_{ij k}$$

where $X_{ij k}$ and $\epsilon_{ij k}$ are the corresponding values of the explanatory variables (i.e. X_{ijk} is a vector) and the error for the k -th unit in the group of farm of type j in region i . Since different fixed effects appear in different groups and the group sizes are small, estimation would be more efficient if it were performed for each parameter on the basis of all the observations involving it jointly. In order to combine fixed effects from different groups in a standard regression analysis an assumption of equal variances is required. This assumption was examined using Bartlett's test for equality of variances in different groups (Bartlett (1947)). The variance estimates taken are the MSE's from the ordinary least squares estimation. The advantage of Bartlett's test is that it can deal with unequal and small group sample sizes. It assumes normality of the error terms, an assumption which seems reasonable upon inspection of residuals from the O.L.S. estimation. Using, therefore, Bartlett's test, it is found that the hypothesis that the variances in the different groups are equal cannot be rejected. Since it appears that equality of variances may be assumed, the above model of fixed effects with equal error variances was applied. According to Table 1 there are six regions and five types of farms. For computational reasons (to prevent multicollinearity), five vectors of coefficients for region effects were introduced into the model, four vectors of coefficients for type of farm effects, and a main effects vector. The coefficients computed using ordinary least squares are presented in Table 4 (a) and (b), where 4(a) is when the dependent variable is production per dunam and 4(b) is when the dependent variable is production value per dunam.

Upon comparing Table 3(a) and (b) (the ordinary least squares within groups estimators) with Table 4(a) and (b) (estimators combining the fixed effects of regions and type of farms) we find that the estimators differ substantially. This may be due to the fact that the o.l.s. estimation within groups is based on very few observations. However, it could also be that in addition to the additive fixed effects there is a random interaction effect, i.e. exchangeability between interaction effects in the regression coefficients. This means that instead of decomposing β_{ij} into

$$\beta_{ij} = \mu + \alpha_i + \gamma_j$$

it is now decomposed into

$$\beta_{ij} = \mu + \alpha_i + \gamma_j + \eta_{ij}$$

with $E(\eta_{ij}) = 0$, i.e. β_{ij} is random. The interpretation in terms of the data is that the explanatory effect of a certain variable which explains productivity can be decomposed into four additive components. The first effect is the basic effect of this variable, the second effect is the additional effect of this variable due to the region the observation is in and the third effect is the additional effect of this variable due to the type of farm the observation is in. The fourth effect (η_{ij}) may be due to a special interaction between the region and type of farm, but since it is not known how to model this effect, we assume it is random, or equivalently we assume an ignorance prior or exchangeability between the interaction terms.

Therefore, we modeled the random interaction effect to be centered around zero with a covariance matrix denoted Δ between the interaction effects of the different variables. The model assumed here can be shown to be equivalent to the model (2). The vector $\underline{\beta}' = (\beta_1, \dots, \beta_r)$ is here $(\underline{\mu}', \underline{\alpha}'_1, \dots, \underline{\alpha}'_I, \underline{\gamma}'_1, \dots, \underline{\gamma}'_J)$ and A_i the matrix of known constants for the i -th group is here A_{ij} for region i and type of farm j . Then estimation proceeds as in §2 and the optimal estimator is given in (9), (10), and (11). However, the quantities σ^2 and Δ are unknown and should be estimated.

Two methods for estimating σ^2 and Δ were proposed, both of which are described in detail in §4. Using both methods the matrix Δ converged to 0, i.e. there appears to be no random effect which can be modeled beyond the fixed effects assumed. So the differences between the regression coefficients in the different groups are due to additive fixed effects.

At this point it was decided to examine these fixed effects more closely for all variables involved in explaining both physical production per dunam and production value per dunam. Several aspects were examined. First it was determined that the main effects for type of farm was not significant. This was determined by an F test comparing a model with a region effect and a type of farm effect to a model with only a region effect ($F^* = 1.53$, $F_{20,120}^{.95} = 1.66$). On the other hand, an F test comparing the model with a region effect and type of farm effect to a model with only a type of farm effect showed that the region effect is significant

($F^* = 1.70$, $F_{24,120}^{.95} = 1.61$). Therefore, the type of farm effect was dropped from the model. Next, we examined whether there is a regional effect on each explanatory variable or whether there is only a difference in intercepts. Again, an F test was performed and only the intercepts were significant ($F^* = 0.95$). Comparing a model with intercepts to a model without intercepts we find that the intercepts are significant and cannot be dropped from the model ($F^* = 10.2$).

At this point it was apparent that the model should include different intercepts for the different regions and only uniform main effects for the explanatory variables. Since the explanatory variables were mostly nonzero it was also decided to re-examine the logarithmic model. Also during execution of these regressions it appeared that the fit to the dependent variable of production value per dunam is better than that to production per dunam (higher R^2 - higher percentage of variance in dependent variable explained by the model). Therefore, both sets of explanatory variables considered thus far for the two dependent variables (see Tables 2, 3, and 4) were combined. It then appeared that the coefficients for area, mechanical equipment, administration costs and depreciation were not significant and these were dropped from the model. Also, we subtracted from the variable of total expense on direct inputs the expense on pesticides, since this expense appeared as a separate variable. The model is then

$$Y_i = \alpha_i + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$$

where $\underline{X}_{1_i}$ is log (age), $\underline{X}_{2_i}$ is log (direct input excluding pesticides) and $\underline{X}_{3_i}$ is log (expense on pesticides). The estimates of the regression coefficients are:

$$\hat{\beta}_1 = .707; \quad \hat{\beta}_2 = .22; \quad \hat{\beta}_3 = .388.$$

TABLE 5 - Region Intercepts

<u>Region</u>	<u>Estimate of Intercept</u>
Inner valeys	1.350
Northern coastal plain	1.130
Yizre'el valley	1.315
Binyamina - Tel Aviv	1.134
Rehovot - Ashqelon	1.000
Lakhish and Negev	1.146

The intercepts are presented in Table 5 and $R^2 = 0.30$. For comparison, $R^2 = 0.15$ for a model without separate intercepts for the different regions. The results for models with the original values of the variables (not logarithms) are similar. The logarithmic model presented may be preferable since it is more easily interpretable as an economic model. The different intercepts for the different regions are viewed as fixed effects and they can also be modelled as a function of some exogeneous variable, i.e.

$$\alpha_i = \alpha + \bar{X}_i' \gamma + \eta_i .$$

In our case, after several runs a model was formed which reduced the number of parameters from six intercepts and three coefficients of variables to one intercept, two coefficients of averages of area and age, and three coefficients of variables. The final proposed model is then

$$Y_{ij} = -1.11 + .564\bar{X}_{1_i} + .17\bar{X}_{2_i} + .431X_{3_{ij}} + .22X_{4_{ij}} + .066X_{5_{ij}}$$

where Y_{ij} is production value per dunam for the j -th unit in the i -th group, $\bar{X}_{1_i}$ is the average age of the orchard in the i -th region, $\bar{X}_{2_i}$ is the average area of the orchard in the i -th region, X_3 is direct input excluding pesticides, X_4 is age and X_5 is expense on pesticides. Note that once the average area of the region was entered, area was no longer a significant explanatory variable. The difference between coefficients of areas in different regions entered the intercepts since the average area of a farm is substantially different in the different regions.

It should be noted that the estimator (13) was not used even for the intercepts since its usefulness is primarily in the case of a large number of groups and several variables. We have in effect only six groups and a covariance matrix of the form $\text{diag}(\delta_0^2, 0, \dots, 0)$ which is not a case for which this estimator is particularly useful.

6. Conclusions

During this project theory and methodology for the regression analysis of data obtained from a complex survey design were developed. The approach presented takes into account the survey design and minimizes expected loss over the whole finite population. Several special cases were examined and problems of covariance components estimation were addressed. The analysis of the data used part of the theory developed.

Both the theory and its applications require further development. Thus for model selection only ad-hoc solutions were available. This problem has solutions for the simple random sample case, but these solutions are not acceptable for more complex designs. In addition, the sample design may possibly be further exploited for estimation purposes. For instance a design-based estimator could replace expressions involving the variances, thus avoiding complicated variance estimation procedures.

The data analysis performed also provided new aspects and prospects. Thus analysis to examine whether the same models hold year after year would be performed. Consecutive data sets could also help test the exchangeability hypothesis. Some variables involve materials which are not put into the soil year after year (e.g. organic fertilizers) and so appear as non-existent in many farms, although they are still in the soil. This affects a model-based analysis and future data sets should include information from the past which might be relevant to the model. Also, all variables presented should be characterized so that they can be translated to fiscal terms. Additional variables, e.g. type of soil, which may be relevant to such an analysis and should be added in the future. Analyzing such data sets on a regular basis might bring more operative results.

The computer programs developed for the EM algorithm, the component-by-component variance estimation and the weighted optimal estimator are fairly general and hopefully will serve many more purposes in the future.

Finally, it should be stressed that the criteria for model selection (i.e. choosing the variables to appear in the model) were purely statistical and not based on any economic theory. We presented a systematic approach to the choice of a model, although this approach was not shown to be optimal in any way. This may have influenced the choice of the final model. We only note that this is the model found given the goal of arriving at a reasonable model within the time and budget limitations.

7. Bibliography

- Bartlett, M.S. (1947). Multivariate Analysis, *Journal of the Royal Statistical Society*, B, 9, 176-197.
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion), *Journal of the Royal Statistical Society*, B, 39, 1-38.
- Dempster, A.P., Rubin, D.B. and Tsutakawa, R.K. (1981). Estimation in Covariance Components Models, *Journal of the American Statistical Association*, 76, 341-353.
- Duncan, D.B. and Horn, S.D. (1972). Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis, *Journal of the American Statistical Association*, 67, 815-822.
- Fay, R.E. III and Herriot, R.A. (1979). Estimation of Income for Small Places: An Application of James-Stein Procedures to Census Data, *Journal of the American Statistical Association*, 74, 269-277.
- Holt, D., Smith, T.M.F. and Winter, P.O. (1980). Regression Analysis of Data from Complex Surveys, *Journal of the Royal Statistical Society*, A, 143, 474-483.
- Hu, T.W. and Stromsdorfer, E.W. (1970). A Problem of Weighting Bias in Estimating the Weighted Regression Model, *Proceedings of the Business and Economic Section of the American Statistical Association*, 513-516.
- Kish, L. and Frankel, M.R. (1974). Inference from Complex Samples (with Discussion), *Journal of the Royal Statistical Society*, B, 36, 1-37.
- Klein, L.R. and Morgan, J. (1951). Results of Alternative Statistical Treatment of Sample Survey Data, *Journal of the American Statistical Association*, 46, 442-460.
- Klein, L.R. (1953). *A Textbook of Econometrics*. Evanston: Row, Peterson & Company.
- Lindley, D.V. and Smith, A.F.M. (1972). Bayes Estimates for the Linear Model (with Discussion), *Journal of the Royal Statistical Society*, B, 34, 1-41.

- Nathan, G. and Holt, D. (1980). The Effect of Survey Design on Regression Analysis, *Journal of the Royal Statistical Society*, B, 42, 377-386.
- Nathan, G. (1981). Regression Analysis Under Differential Non-Response, *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Nathan, G. (1983). A simulation Comparison of Estimators for a Regression Coefficient Under Differential Non-Response, *Communications in Statistics: Theory and Methods*, 11 (to appear).
- Pfeffermann, D. (1978). Regression Analysis for Complex Samples from Finite Populations. Unpublished Ph.D. Thesis, Hebrew University, Jerusalem.
- Pfeffermann, D. and Nathan, G. (1981). Regression Analysis of Data from a Cluster Sample, *Journal of the American Statistical Association*, 76, 681-689.
- Pfeffermann, D. (1982). Multivariate Prediction on Finite Population Sampling with Missing Data. Paper presented at the International Meeting on Analysis of Sample Survey Data and Sequential Analysis, Jerusalem.

Description and Results of Cooperation

Cooperation worked in two directions. First, the Statistical Analysis Unit of the Central Bureau of Statistics worked in close contact with Mr. Gad Bental of the Institute of Farm Income Research. Mr. Bental was involved in all steps of the analysis via meetings at the Central Bureau of Statistics. His advice and explanations helped the Bureau team to better understand both the underlying agricultural mechanisms and economic aspects. Mr. Bental was consulted also on the reasonability of various assumptions in the model with respect to the citrus data.

The second direction of cooperation is with Prof. W. Fuller of Iowa State University. There was some correspondence with Prof. Fuller, but his main contribution was when he was in Israel in June 1982. Prof. Fuller read all that was done, commented and improved on the existing material, offering suggestions from his vast experience in statistics applied to agricultural studies. Further input was obtained from Prof. Fuller and from others at Iowa State University in the course of a visit of one of the senior investigators in the U.S. during November 1982. In particular, the fixed additive effects model for the regression coefficients and the testing of a random effects interaction term were introduced on the basis of doubts expressed with respect to the assumption of exchangeability.

OMED1 - A program to calculate $\beta_i(a)$ $i = 1, \dots, n$ and $\beta(a)$ defined by (3) and (4), pages 4, 5. The parameters Δ and σ^2 are input to the program and not computed.

```

ATTACH(INSLEFTN)
LIBRARY(INSLEFTN)
ATTACH(TAPE1, GRAPE22, ID=KTKH, MR=1)
ATTACH(TAPE2, DESM22, ID=KTKH, MR=1)
ATTACH(TAPE8, DEL14, ID=KTKH, MR=1)
ATTACH(TAPE9, AMAT1, ID=KTKH, MR=1)
REQUEST(8, *PF)
FTN(OPT=0, PMD)
RFL(200000)
LGO.
REDUCE.
REWIND(TAPE11)
COPYUR(TAPE11, 8, EOI)
CATALOG(8, BETA, ID=KTKH)
EXIT(U)
REWIND(ZZZZZMP)
COPYSBF(ZZZZZMP)
PROGRAM OMED(INPUT, OUTPUT, TAPE2, TAPE8=65, TAPE9, TAPE11)
COMMON DEL1(4,4), C(40,40), AD(40,40),
*ADA(40,40), DA(4,40), DAC(4,40), D(22,4,4), D11(88,88), B(4,4), S(4),
*E(88), D22(88,88), DE(88,88), BETA(88)
INTEGER IJOB, N, IA, IER, R, A(4,40), A1(22,4,40)
REAL D1, D2, WKAREA(220), BE(1)
N=4
IA=4
IJOB=1
D1=-1.0
KK=22
SIGSQ=2.8261
R=40
NK=88

DO 5 I=1, N
5 READ(8, 77) (DEL1(I, J), J=1, N)
77 FORMAT(2X, 5F12.4)
DO 6 I=1, N
6 PRINT 77, (DEL1(I, J), J=1, N)
CALL LINV3F(DEL1, BE, IJOB, N, IA, D1, D2, WKAREA, IER)
PRINT *, IER
DO 7 I=1, N
7 PRINT 77, (DEL1(I, J), J=1, N)
DO 10 I=1, R
DO 10 J=1, R
10 C(I, J)=0.0

DO 15 NMAT=1, KK
DO 20 I=1, N
20 READ(9, 88) (A(I, J), J=1, R)
88 FORMAT(2X, 50I2)
DO 22 I=1, N
DO 22 J=1, R
22 A1(NMAT, I, J)=A(I, J)
DO 30 I=1, R
DO 30 J=1, N
AD(I, J)=0.0
DO 30 K=1, N
30 AD(I, J)=AD(I, J)+A(K, I)*DEL1(K, J)

DO 40 I=1, R
DO 40 J=1, R
ADA(I, J)=0.0
DO 40 K=1, N
40 ADA(I, J)=ADA(I, J)+AD(I, K)*A(K, J)
DO 50 I=1, R
DO 50 J=1, R
50 C(I, J)=C(I, J)+ADA(I, J)
15 CONTINUE

CALL LINV3F(C, BE, IJOB, R, R, D1, D2, WKAREA, IER)
PRINT *, IER
DO 60 I=1, R
60 PRINT 78, (C(I, J), J=1, R)
78 FORMAT((2X, 10F13.4))

DO 70 NM=1, KK
I1=4*NM-3
I2=4*NM
DO 80 J=1, N

```

```

DO 80 J=1,R
DA(I,J)=0.0
DO 80 K=1,N
80 DA(I,J)=DA(I,J)+DEL1(I,K)*A1(NM,K,J)
PRINT *,NM,I1,I2
DO 90 J=1,N
90 PRINT 78,(DA(I,J),J=1,R)

```

```

DO 100 I=1,N
DO 100 J=1,R
DAC(I,J)=0.0
DO 100 K=1,R
100 DAC(I,J)=DAC(I,J)+DA(I,K)*C(K,J)

DO 130 NMAT=1,KK
DO 140 I=1,R
DO 140 J=1,N
AD(I,J)=0.0
DO 140 K=1,N
140 AD(I,J)=AD(I,J)+A1(NMAT,K,I)*DEL1(K,J)

```

```

DO 150 I=1,N
DO 150 J=1,N
D(NMAT,I,J)=0.0
DO 150 K=1,R
150 D(NMAT,I,J)=D(NMAT,I,J)+DAC(I,K)*AD(K,J)
PRINT *,NM,NMAT
DO 155 I=1,N
155 PRINT 77,(D(NMAT,I,J),J=1,N)
J1=4*NMAT-3
J2=4*NMAT

```

```

DO 160 I=I1,I2
DO 160 J=J1,J2
L1=I-(NM-1)*4
L2=J-(NMAT-1)*4
160 D11(I,J)=D(NMAT,L1,L2)
130 CONTINUE
70 CONTINUE

```

```

DO 170 I=1,NK
DO 170 J=1,NK
170 D22(I,J)=0.0

```

```

DO 180 NMAT=1,KK
DO 185 I=1,N
185 READ(2,89) (S(I),(B(I,J),J=1,N))
89 FORMAT(2X,6F10.2)
M1=4*NMAT-3
M2=4*NMAT
DO 190 I=M1,M2
L1=I-(NMAT-1)*4
E(I)=S(L1)/SIGSQ
DO 190 J=M1,M2
L2=J-(NMAT-1)*4
190 D22(I,J)=B(L1,L2)/SIGSQ
180 CONTINUE

```

```

DO 200 I=1,NK
DO 200 J=1,NK
200 DE(I,J)=D11(I,J)+D22(I,J)
PRINT 78,(E(I),I=1,NK)
CALL LINV3F(DE,BE,IJOB,NK,NK,D1,D2,WKAREA,IER)
PRINT *,IER

```

```

DO 210 I=1,NK
BETA(I)=0.0
DO 210 K=1,NK
210 BETA(I)=BETA(I)+DE(I,K)*E(K)
PRINT 78,(BETA(I),I=1,NK)
WRITE(11,76) (BETA(I),I=1,NK)
76 FORMAT((2X,4F13.4))
STOP
END

```

OMED2 - A program to calculate the estimate of B^*Y_{-s} defined by (9),

page 7.

```

ATTACH(IMSLEFTN)
LIBRARY(IMSLEFTN)
ATTACH(TAPE1, GRAPE22, ID=KTKH, MR=1)
ATTACH(TAPE10, BETA, ID=KTKH, MR=1)
FTN(OPT=0, PMD)
LGO.
EXIT(U)
REWIND(ZZZZZZMP)
COPYSBF(ZZZZZZMP)
PROGRAM OMED2(INPUT, OUTPUT, TAPE1, TAPE10)
DIMENSION X(28,4), Y(28), W(28), BETA(4), NROW(22), SUMXWB(4),
*SUMXW(4,4), SUMXX(4,4), XX(28,4,4), XWB(4), DY(4)
INTEGER IJOB, N, IA, IER
REAL D1, D2, WKAREA(10), BE(1)
DATA NROW/28, 11, 11, 8, 10, 6, 2, 3, 5, 13, 13, 2, 11, 17, 1, 2, 3, 7, 15, 10, 6, 3/
N=4
IA=4
IJOB=1
D1=-1
KK=22
DO 5 I=1, N
SUMXWB(I)=0.0
DO 5 J=1, N
5 SUMXW(I, J)=0.0

DO 15 NMAT=1, KK
MN=NROW(NMAT)
DO 20 I=1, MN
20 READ(1, 99) (Y(I), (X(I, J), J=1, N), W(I))
99 FORMAT(14X, F7.4, F5.2, 7X, F7.4, 14X, 2F7.4, 55X, F7.2)
READ(10, 76) (BETA(I), I=1, N)
76 FORMAT(2X, 4F13.4)
DO 25 I=1, N
DO 25 J=1, N
25 SUMXX(I, J)=0.0
DO 30 M=1, MN
DO 40 I=1, N
DO 40 J=1, N
XX(M, I, J)=X(M, I)*Y(M, J)*W(M)
40 SUMXX(I, J)=SUMXX(I, J)+XX(M, I, J)
30 CONTINUE
PRINT *, NMAT, MN
DO 50 I=1, N
50 PRINT 77, (SUMXX(I, J), J=1, N)
77 FORMAT(2X, 5F12.4)

DO 60 I=1, N
XWB(I)=0.0
DO 60 K=1, N
60 XWB(I)=XWB(I)+SUMXX(I, J)*BETA(I)
DO 70 I=1, N
70 SUMXWB(I)=SUMXWB(I)+XWB(I)
PRINT *, NMAT
PRINT 77, (SUMXWB(I), I=1, N)
DO 80 I=1, N
DO 80 J=1, N
80 SUMXW(I, J)=SUMXW(I, J)+SUMXX(I, J)
DO 85 I=1, N
85 PRINT 77, (SUMXW(I, J), J=1, N)
15 CONTINUE

CALL LINV3F(SUMXW, BE, IJOB, N, IA, D1, D2, WKAREA, IER)
PRINT *, IER
DO 90 I=1, N
90 PRINT 77, (SUMXW(I, J), J=1, N)

DO 100 I=1, N
DY(I)=0.0
DO 100 K=1, N
100 DY(I)=DY(I)+SUMXW(I, K)*SUMXWB(K)
PRINT 888
888 FORMAT(3X, *OPTIMAL ESTIMATOR DY(S)*)
PRINT 77, (DY(I), I=1, N)
STOP
END

```


NV100 - A program to calculate the estimate of Δ (Pfeffermann (1978))

by the component by component approach. The parameter σ^2 is an input

to the program.

```

ATTACH(IPLFTN)
LIBRARY(IMSLEFTH)
ATTACH(TAPE2,DES5100,IO=KTKH,MR=1)
ATTACH(TAPE5,DEL100,IO=KTKH,MR=1)
ATTACH(TAPE9,AMAT515,IO=KTKH,MR=1)
FTN(OPT=0,PMD,PL=6000)
LGO.

PROGRAM NAIVE(INPUT,OUTPUT,TAPE2,TAPE4,TAPE5,TAPE6,TAPE7,TAPE8,
*TAPE9,TAPE11)
DIMENSION S(5),H(5,5),BS(5),DELZ(5,5),RSM(15,5),RT(15,5,5),
*BFT(10),SUMO(10),APET(5),DELI(5,5),AB(15,5),DELUP(5,5),LEL(10,5,5),
*DELEL(5,5),RINV(5,5),AR(5,5),PSYM(15)
REAL C1,C2,WKAREA(200),RE(1),LAMBDA(10),LAMA(10,10),LAM(5),
*LM(15,5),LAMINV(10,10),LE(10,10),LSYM(55)
INTEGER IJOB,N,IA,IER,R,A(5,50),T
N=5
IA=5
IJOB=1
C1=-1.0
KK=15
SIGZ=2.2005
M=5
IJ=1
R=50
T=10

1 DO 10 I=1,N
10 READ(5,77) (DELZ(I,J),J=1,N)
77 FORMAT(2X,5F12.4)
DO 15 I=1,N
15 PRINT 77,(DELZ(I,J),J=1,N)
IF(IJ.EQ.1) GO TO 22

C--- THIS SECTION IS RUN ONLY FROM SECOND ITERATION
C---
C9 REWIND 6
REWIND 7
REWIND 9
PRINT *,M,IJ
PRINT *,DIF,DELI(M,M)
DELEL(M,M)=DELI(M,M)
PRINT *,DELEL(M,M)
IJ=IJ+1

C--- COMPUTING LAMDA AND BETA
C---
22 IF(M.GT.5.OR.IJ.GT.1) 41,24
M=2
24 DO 30 NMAT=1,KK
DO 35 I=1,N
35 READ(2,88) (S(I),(B(I,J),J=1,N))
K=1
DO 25 I=1,N
C 25 PRINT 77,(R(I,J),J=1,N)
DO 29 J=1,N
DO 29 J=1,I
BSYM(K)=B(I,J)
29 K=K+1
CALL LINV3P(RSYM,RE,IJOB,N)
PRINT *,IER,NMAT
C 25 PRINT 77,(BSYM(K),K=1,5)
C 25 PRINT 77,(RSYM(K),K=6,10)
C 25 PRINT 77,(RSYM(K),K=11,15)
68 FORMAT(2X,6F10.2)
K=1
DO 23 I=1,N
DO 23 J=1,I
BINV(I,J)=BSYM(K)
23 K=K+1
DO 31 I=1,N
DO 31 J=1,N
31 BINV(I,J)=BINV(J,I)
C 26 PRINT 77,(BINV(I,J),J=1,N)
C 26 CALL VMULFF(B,BINV,IA,IA,IA,IA,IA,IA,IA,IER)
C 27 DO 27 I=1,N
C 27 PRINT 77,(R(I,J),J=1,N)
DO 28 I=1,N

```

```

28 DO 28 J=1,N
   B(1,J)=RINV(1,J)
   DO 40 I=1,N
     BS(I)=C.0
   DO 40 K=1,N
   40 BS(I)=BS(I)+B(I,K)*S(K)
     PRINT 78,((RS(I),I=1,N),(B(1,I),I=1,N))
     WRITE(4,78)((BS(I),I=1,N),(B(1,I),I=1,N))
     WRITE(8,78)(B(1,2),B(1,3),B(1,4),B(1,5),B(2,3),B(2,4),B(2,5),
      *B(3,4),B(3,5),B(4,5))
78 FORMAT(2X,10F12.4)
30 CONTINUE

```

```

41 REWIND 4
   DO 43 I=1,T
     LAMBA(I)=0.0
   DO 43 J=1,T
43 LAMA(I,J)=C.0

```

```

   DO 45 NMAT=1,KK
     READ(4,78)((RS(I),I=1,N),(B(1,I),I=1,N))
     PRINT *,DELZ(M,M),SIGZ,B(M,M)
     LAM(M)=1/(DELZ(M,M)+SIGZ*B(M,M))
     DO 50 I=1,N
50 READ(9,87)(A(I,J),J=1,P)
87 FORMAT(2X,50I2)
     PRINT *,NMAT
     PRINT 79,(LAM(M),(A(M,J),J=1,P))
     WRITE(6,79)(LAM(M),(A(M,J),J=1,P))
79 FORMAT(2X,F12.4,2X,50I2)
     DO 60 I=1,T
       L2=M+(I-1)*5
     DO 60 J=1,T
       L1=M+(J-1)*5
60 LAMA(I,J)=LAMA(1,J)+LAM(M)*A(M,L1)+A(M,L2)
   DO 62 I=1,T
C 62 PRINT 78,(LAMA(1,J),J=1,T)
   DO 70 I=1,T
     L2=M+(I-1)*5
70 LAMBA(I)=LAMBA(1)+LAM(M)*BS(M)*A(M,L2)
45 CONTINUE
   DO 64 I=1,T
64 PRINT 78,(LAMA(I,J),J=1,T)

```

C--- COMPUTING BETAGAL AND DELTA1
C---

```

   K=1
   DO 71 I=1,T
     DO 71 J=1,T
       LSYM(K)=LAMA(I,J)
71 K=K+1
   CALL LINV3P(LSYM,LE,IJCB,T,IER)
   PRINT *,IER
   K=1
   DO 72 I=1,T
     DO 72 J=1,T
       LAMINV(I,J)=LSYM(K)
72 K=K+1
   DO 73 I=1,T
     DO 73 J=1,T
73 LAMINV(I,J)=LAMINV(J,I)
   DO 75 I=1,T
75 PRINT 78,(LAMINV(I,J),J=1,T)
   CALL VMULFF(LAMA,LAMINV,I,T,T,T,T,LL,T,IER)
   DO 76 I=1,T
76 PRINT 78,(LL(I,J),J=1,T)
   DO 81 I=1,T
     DO 81 J=1,T
81 LAMA(I,J)=LAMINV(I,J)
   DO 80 I=1,T
     DO 80 K=1,T
80 BET(I)=BET(I)+LAMA(I,K)*LAMBA(K)
   PRINT 78,(BET(I),I=1,T)

```

```

REWIND 4-
REWIND 6
SUMD(M)=0.0
DO 100 NMAT=1,KK
  READ(4,77)(RS(I),I=1,N)
  READ(6,79)(LAM(M),(A(M,J),J=1,P))
  ABET(M)=C.0

```

```

DO 105 I=1,I
  L1=M+(I-1)*5
105 ABET(M)=ABET(M)+A(M,L1)*RET(I)
  WRITE (7,101) ABET(M)
  PRINT 101,ABET(M)
101 FORMAT(2X,F12.4)
  SUMD(M)=SUMD(M)+LAM(M)*(BS(M)-ABET(M))*2
100 CONTINUE
  DEL1(M,M)=(DELZ(M,M)/(KK-1))*SUMD(M)

```

C-- CHECKING CONDITIONS FOR CONVERGENCE

```

C--
  DIF=(DEL1(M,M)-DEL7(M,M))/DEL1(M,M)
  IF(ABS(DIF).GT.0.001) 99,109
109 PRINT *,M,IJ
  PRINT *,DIF,DEL1(M,M)
  REWIND 6
  REWIND 7
  DO 120 NMAT=1,KK
    READ(6,101) LAM(M)
    READ(7,101) ABET(M)
    WRITE(11,77) DEL1(M,M),LAM(M),ABET(M)
120 CONTINUE

```

```

  REWIND 5
  REWIND 6
  REWIND 7
  REWIND 9
  M=M+1
  IJ=1
  IF(M.LE.5) 1,149

```

C--- COMPUTING COVARIANCE DEL(L,L')

```

C---
149 REWIND 4
  REWIND 8
  REWIND 11
  DO 125 I=1,N
    DO 130 NMAT=1,KK
      IF(I.GT.1) GO TO 143
      READ(8,78) (B(1,2),B(1,3),B(1,4),B(1,5),B(2,3),B(2,4),B(2,5),
        *B(3,4),B(3,5),B(4,5))
      READ(4,76) (BS(L),L=1,N)
      DO 140 L=1,N
        BSM(NMAT,L)=BS(L)
        BI(NMAT,1,2)=B(1,2)
        BI(NMAT,1,3)=B(1,3)
        BI(NMAT,1,4)=B(1,4)
        BI(NMAT,1,5)=B(1,5)
        BI(NMAT,2,3)=B(2,3)
        BI(NMAT,2,4)=B(2,4)
        BI(NMAT,2,5)=B(2,5)
        BI(NMAT,3,4)=B(3,4)
        BI(NMAT,3,5)=B(3,5)
        BI(NMAT,4,5)=B(4,5)
143 READ(11,77) DEL1(I,I),LAM(I),ABET(I)
        LM(NMAT,I)=LAM(I)
        AB(NMAT,I)=ABET(I)
130 CONTINUE
125 CONTINUE

```

```

  PRINT *,M
  DO 132 I=1,KK
132 PRINT 77,(BSM(I,J),J=1,N)
  PRINT *,M
  DO 134 I=1,KK
134 PRINT 77,(LM(I,J),J=1,N)
  PRINT *,M
  DO 136 I=1,KK
136 PRINT 77,(AB(I,J),J=1,N)
  PRINT *,M
  PRINT 77,(DEL1(I,I),I=1,N)

```

```

  L=4
  DO 150 I=1,N
  DO 150 J=1,N
    DELUP(I,J)=0.0
    DELDO(I,J)=0.0

```

150 DELEL(I,J)=0.0

- 50 -

DO 155 NMAT=1, KK

DO 160 I1=1, L

I=I1+1

DO 160 I2=1, N

DELD0(I1,I2)=DELD0(I1,I2)+SQR1(LM(NMAT,I1))*SQRT(LM(NMAT,I2))

DELUP(I1,I2)=DELUP(I1,I2)+SQR1(LM(NMAT,I1))*SQRT(LM(NMAT,I2))*

((BSM(NMAT,I1)-AB(NMAT,I1))(PSM(NMAT,I2)-AB(NMAT,I2))-S1C7*

*BI(NMAT,I1,I2))

PRINT *, I1, I2, NMAT, DELUP(I1,I2), DELD0(I1,I2)

160 CONTINUE

155 CONTINUE

DO 170 I1=1, L

I=I1+1

DO 170 I2=1, N

DELEL(I1,I2)=DELUP(I1,I2)/(KK*DELD0(I1,I2))

PRINT *, I1, I2, DELEL(I1,I2)

170 CONTINUE

DO 180 I=1, N

DO 180 J=1, I

IF(I.EQ.J)-183,185

183 DELEL(I,J)=DELEL(I,J)

GO TO 180

185 DELEL(I,J)=DELEL(J,I)

180 CONTINUE

PRINT 777

777 FORMAT(5X, *VARIANCE COVARIANCE MATRIX DELTA(N,N)*//)

DO 190 I=1, N

190 PRINT 77, (DELEL(I,J), J=1, N)

STOP

END

EMI - A program to calculate the estimates of Δ and σ^2 using the EM algorithm. (Dempster, Rubin, Tsutakawa (1981)).

```

ATTACH(TAPE1,GR5100,ID=KTKH,MP=1)
LIBRARY(IMS1FTN)
ATTACH(TAPE2,DES5100,ID=KTKH,MR=1)
ATTACH(TAPE3,DEL100,ID=KTKH,MF=1)
ATTACH(TAPE4,AMAT515,ID=KTKH,MP=1)
RECLEST(N,*PF)
FTN(OPT=C,PMD,PL=5000)
LGD.
REWIND(TAPE8)
CCPYUP(TAPE8,N,,EOI)
CATALOG(N,DEL1A,ID=KTKH)
REWIND(N)
COPYS8F(N)
PRCGRAM EM (INPUT,OUTPUT,TAPE1,TAPE2,TAPE3,TAPE4,TAPE5,TAPE6,TAPE7,TAPE8,
*TAPE9)
DIMENSION DEL(5,5),SIG(1),SUMAS(50),SUMABDS(50),SUMABAS(50,50),
*SABDBA(50,50),S(5),R(5),D(5,5),FU(5,5),FDR(5,5),AR(50,5),
*ABA(50,50),ABDB(50,5),AHDBA(50,50),RDS(5),ARDS(50),AS(50),
*H(50,50),HAS(50),HARDS(50),UC(50),CR(5,5),DS(5,5),DAS(5,50),
*EL2(5),EL3(5),BETA(5),SUMBD(5,5),SUMBD40(5,5),SUMBD60(5,5),
*CELL(5,5),SUMABAH(50,50),PETA(50),PETA9(5),SA(50),RA(5,50),
*ABAH(50,50),HA(50,5),BAHA(5,5),BAHARD(5,5),DBAHARD(5,5),PBD(5,5),
*BFTRFT(5,5),DC(5,5),UARA(50),SOG(5,5),RLOG(5,5),RSQRB(5,5),
*HINV(50,50),BINV(5,5),FH(50,50),RE(5,5),BSYM(15),HSYM(127),
*SGINV(50,5),SGSYM(15),SGSG(5,5),DASABDS(50),YMD(50),SUMDIF(50)
INTEGER IJDB,N,IA,IER,NPDW(15),R,A(5,50)
REAL D1,D2,KKAREA(2750),RE(1)
DATA NPCW/28,11,11,9,10,6,13,13,11,17,7,7,15,10,6/
N=5
KK=15
TA=5
IJDB=1
D1=-1.0
SIG7=2.28074
IJ=1
R=50

DO 1 I=1,N
  1 READ(5,77) (DEL(I,J),J=1,N)
  77 FORMAT(2X,5F12.4)
  DO 4 J=1,N
    4 PRINT 77,(DEL(I,J),J=1,N)
    IF(IJ,EOI) GO TO 14
C--- THIS SECTION IS RUN ONLY FROM SECOND ITERATION
C---
  99 REWIND 1
  REWIND 2
  REWIND 6
  REWIND 7
  REWIND 9
  IJ=IJ+1
  SIG7=SIG(IJ-1)
  PRINT *,SIG7
  DO 12 I=1,N
    DO 12 J=1,N
      12 DEL(I,J)=DEL1(I,J)
    DO 13 I=1,N
      13 PRINT 77,(DEL(I,J),J=1,N)

C--- COMPUTATIONS FOR 3 FIRST STEPS OF ITERATIVE PROCEDURE
C---
  14 DO 8 J=1,R
    SUMAS(I)=SUMABDS(I)=SUMDIF(I)=0.0
    DO 8 J=1,R
      8 SUMABA(I,J)=SABDBA(I,J)=0.0

  DO 10 NMAT=1,KK
    DO 11 I=1,N
      11 READ(2,88) (S(I),(P(I,J),J=1,N))
      88 FORMAT(2X,6F10.2)
      PRINT 112,NMAT
      PRINT 88,(S(I),I=1,5)
      112 FORMAT(2X,*I TIMES XI *,13)
      DO 111 I=1,N
        C 111 PRINT 77,(B(I,J),J=1,N)
      K=1

```

```

DO 115 I=1,N
DO 115 J=1,N
BSYM(K)=B(I,J)
115 K=K+1
CALL LINV3P(BSYM,85,IJOB,N,IER)
PRINT *,IER,NMAT
PRINT 77,(BSYM(K),K=1,5)
PRINT 77,(BSYM(K),K=6,10)
PRINT 77,(BSYM(K),K=11,15)
K=1
DO 116 I=1,N
DO 116 J=1,N
B(I,J)=BSYM(K)
116 K=K+1
DO 117 I=1,N
DO 117 J=1,N
117 B(I,J)=B(J,I)
DO 15 I=1,N
15 PRINT 77,(BINV(I,J),J=1,N)
CALL VMULFF(B,BINV,IA,IA,IA,IA,IA,88,IA,IER)
DO 16 I=1,N
16 PRINT 77,(AB(I,J),J=1,N)
DO 17 I=1,N
DO 17 J=1,N
17 B(I,J)=PINV(I,J)
DO 18 I=1,N
DO 18 J=1,N
18 SOG(I,J)=B(I,J)+(1.0/SIGZ)*DEL(I,J)
PRINT 21
21 FORMAT(2X,*X+X+1/S TIMES DELTA*)
DO 19 I=1,N
19 PRINT 77,(SOG(I,J),J=1,N)
K=1
DO 26 I=1,N
DO 26 J=1,N
SGSYM(K)=SOG(I,J)
26 K=K+1
CALL LINV3P(SGSYM,85,IJOB,N,IER)
PRINT *,IER
PRINT 77,(SGSYM(K),K=1,5)
PRINT 77,(SGSYM(K),K=6,10)
PRINT 77,(SGSYM(K),K=11,15)
K=1
DO 28 I=1,N
DO 28 J=1,N
SOG(I,J)=SGSYM(K)
28 K=K+1
DO 31 I=1,N
DO 31 J=1,N
31 SOG(I,J)=SOG(J,I)
DO 20 I=1,N
20 PRINT 77,(SOG(I,J),J=1,N)
CALL VMULFF(SOG,SGINV,IA,IA,IA,IA,IA,SGSG,IA,IER)
DO 22 I=1,N
22 PRINT 77,(SGSG(I,J),J=1,N)
DO 24 I=1,N
DO 24 J=1,N
24 SOG(I,J)=SGINV(I,J)

DO 23 I=1,N
DO 23 J=1,N
BSOG(I,J)=0.0
DO 23 K=1,N
23 BSOG(I,J)=BSOG(I,J)+B(I,K)*SOG(K,J)
DO 25 I=1,N
DO 25 J=1,N
BSOGB(I,J)=0.0
DO 25 K=1,N
25 BSOGB(I,J)=BSOGB(I,J)+BSOG(I,K)*B(K,J)
DO 27 I=1,N
DO 27 J=1,N
27 C(I,J)=B(I,J)-BSOGB(I,J)
DO 29 I=1,N
PRINT 77,(C(I,J),J=1,N)
29 WRITE(6,77) (C(I,J),J=1,N)
10 CONTINUE

```

```

REWIND 2
REWIND 4
DO 30 NMAT=1,KK
DO 32 I=1,N

```

```

32 READ(2,86) (S(I),(B(I,J),J=1,N))
DO 34 I=1,N
34 READ(9,87) (A(I,J),J=1,P)
87 FORMAT(2X,50I2)
DO 36 I=1,N
36 READ(6,77) (D(I,J),J=1,N)
DO 37 I=1,N
DO 37 J=1,N
PD(I,J)=0.0
DO 37 K=1,N
37 BD(I,J)=BD(I,J)+B(I,K)*D(K,J)
DO 40 I=1,N
DO 40 J=1,N
BDB(I,J)=0.0
DO 40 K=1,N
40 BDB(I,J)=BDB(I,J)+BD(I,K)*B(K,J)
PRINT 41,NMAT
41 FORMAT(5X,* BDB *,15)
DO 43 I=1,N
43 PRINT 77,(RDB(I,J),J=1,N)

```

```

DO 42 I=1,R
DO 42 J=1,N
AB(I,J)=0.0
DO 42 K=1,N
42 AB(I,J)=AB(I,J)+A(K,I)*R(K,J)
DO 45 I=1,R
DO 45 J=1,P
ARA(I,J)=0.0
DO 45 K=1,N
45 ABA(I,J)=ABA(I,J)+AB(I,K)*A(K,J)
PRINT 44
44 FORMAT(5X,*ABA *)
DO 46 I=1,N
46 PRINT 77,(ABA(I,J),J=1,N)
DO 47 I=1,R
DO 47 J=1,R
47 SUMABA(I,J)=SUMABA(I,J)+ABA(I,J)

```

```

DO 50 I=1,R
DO 50 J=1,N
ABDB(I,J)=0.0
DO 50 K=1,N
50 ABDB(I,J)=ABDB(I,J)+A(K,I)*BDB(K,J)
DO 52 I=1,R
DO 52 J=1,R
ABDPA(I,J)=0.0
DO 52 K=1,N
52 ABDPA(I,J)=ABDPA(I,J)+ABDB(I,K)*A(K,J)
PRINT 51
51 FORMAT(5X,*ABDPA *)
DO 53 I=1,R
53 PRINT 78,(ABDPA(I,J),J=1,R)
DO 55 I=1,R
DO 55 J=1,R
55 SABDPA(I,J)=SABDPA(I,J)+ABDPA(I,J)

```

```

DO 60 I=1,N
BDS(I)=0.0
DO 60 K=1,N
60 BDS(I)=BDS(I)+BD(I,K)*S(K)
DO 62 I=1,R
APDS(I)=0.0
DO 62 K=1,N
62 ABDS(I)=ABDS(I)+A(K,I)*BDS(K)
PRINT 61
61 FORMAT(5X,*ABDS FIRST N TERMS *)
PRINT 79,(ABDS(I),I=1,N)
DO 64 I=1,R
64 SUMABDS(I)=SUMABDS(I)+ABDS(I)
DO 66 I=1,R
AS(I)=0.0
DO 66 K=1,N
66 AS(I)=AS(I)+A(K,I)*S(K)
PRINT 67
67 FORMAT(5X,*AS *)
PRINT 78,(AS(I),I=1,N)
DO 69 I=1,P
DASABDS(I)=AS(I)-APDS(I)
69 SUMDIF(I)=SUMDIF(I)+DASABDS(I)
PRINT 78,(DASABDS(I),I=1,P)
DO 68 I=1,R

```

```

68 SUMAS(I)=SUMAS(I)+AS(I)
PRINT *,NMAT
78 FORMAT(2X,10F13.4)
30 CONTINUE
PRINT 78,(SUMABDS(I),I=1,R)
PRINT 78,(SUMAS(I),I=1,R)

```

C--- COMPUTING POSTERIOR ESTIMATES FOR UO,BETA1(O) TO BETA(N(C))
C---

```

DO 70 I=1,R
DO 70 J=1,R
70 H(I,J)=SUMABA(I,J)-SABDBA(I,J)
DO 699 I=1,R
C 699 PRINT 78,(H(I,J),J=1,R)
K=1
DO 71 I=1,R
DO 71 J=1,I
HSYM(K)=H(I,J)
71 K=K+1
CALL LINV3P(HSYM,HE,IJOB,R,1FR)
PRINT *,IER
PRINT 888
888 FORMAT(5X,*INVERSE OF H MATRIX*//)
K=1
DO 72 I=1,R
DO 72 J=1,I
H(I,J)=HSYM(K)
72 K=K+1
DO 73 I=1,R
DO 73 J=1,R
73 H(I,J)=H(J,I)
DO 74 I=1,R
74 PRINT 78,(HINV(I,J),J=1,R)
CALL VMULFF(H,HINV,R,R,R,R,R,HH,R,IER)
DO 76 I=1,R
76 PRINT 77,(HH(I,J),J=1,R)
DO 79 I=1,R
DO 79 J=1,R
79 H(I,J)=HINV(I,J)
DO 82 I=1,R
XMU(I)=0.0
DO 82 J=1,R
82 XMU(I)=XMU(I)+H(I,J)*SUMDIF(J)
PRINT 83
83 FORMAT(5X,*NEW MU*)
PRINT 78,(XMU(I),I=1,R)
DO 75 I=1,R
HAS(I)=0.0
HABDS(I)=0.0
DO 75 K=1,R
HAS(I)=HAS(I)+H(I,K)*SUMAS(K)
75 HABDS(I)=HABDS(I)+H(I,K)*SUMABDS(K)
PRINT 78,(HAS(I),I=1,R)
PRINT 78,(HABDS(I),I=1,R)
DO 80 I=1,R
80 UO(I)=HAS(I)-HABDS(I)
PRINT 81
81 FORMAT(5X,*MU-ZERO*)
PRINT 78,(UO(I),I=1,R)

```

```

REWIND 2
REWIND 6
REWIND 9
DO 100 NMAT=1, KK
DO 100 I=1, N
READ(2,86) (S(I),(R(I,J),J=1,N))
READ(9,87) (A(I,J),J=1,N)
100 READ(6,77) (D(I,J),J=1,N)
DO 110 I=1,R
AS(I)=0.0
DO 110 K=1,N
110 AS(I)=AS(I)+A(K,I)*S(K)
PRINT 113,NMAT
113 FORMAT(5X,*A1 TIMES S1*,15)
PRINT 78,(AS(K),K=1,N)
DO 120 I=1,N
DO 120 J=1,N
DB(I,J)=0.0
DO 120 K=1,N
120 DB(I,J)=D9(I,J)+D(I,K)*R(K,J)
DO 121 I=1,N

```



```

DO 122 I=1,N
DO 122 J=1,N
DBA(I,J)=0.0
DO 122 K=1,N
122 DBA(I,J)=DBA(I,J)+DB(I,K)*A(K,J)
C 123 PRINT 78, (DBA(I,K), K=1,50)
DO 124 I=1,N
DS(I)=0.0
DO 124 K=1,N
124 DS(I)=DS(I)+D(I,K)*S(K)
PRINT 77, (DS(K), K=1,N)
DO 127 I=1,N
EL2(I)=0.0
DO 127 K=1,N
127 EL2(I)=EL2(I)+DBA(I,K)*HAS(K)
PRINT 77, (EL2(I), I=1,N)
DO 130 I=1,N
EL3(I)=0.0
DO 130 K=1,N
130 EL3(I)=EL3(I)+DBA(I,K)*HABDS(K)
PRINT 77, (EL3(I), I=1,N)
DO 140 I=1,N
140 BETA(I)=DS(I)-EL2(I)+EL3(I)
PRINT *, NMAT
PRINT 77, (BETA(I), I=1,N)
WRITE(7,77) (BETA(I), I=1,N)
100 CONTINUE

```

C--- COMPUTING SIGMA SQUARE FOR EACH ITERATION (DEFINED BY SIG(I,J))
C---

```

SUMYS=0.0
141 READ(1,98) Y0
98 FORMAT(14X,F7.4)
IF (EOF(1)) 147,145
145 SUMYS=SUMYS+Y0**2
GO TO 141
147 PRINT *, SUMYS
REWIND 2
REWIND 6
REWIND 7
REWIND 9

```

```

SBETA=BETABAU=BETBBET=SAU=UABAU=0.0
DO 175 I=1,N
DO 175 J=1,N
175 SUMBD(I,J)=SUMB4D(I,J)+SUMB6D(I,J)+DEL1(I,J)=0.0
DO 176 I=1,N
DO 176 J=1,N
176 SUMABAH(I,J)=0.0
SUM=0.0

```

```

DO 180 NMAT=1,KK
MN=NROW(NMAT)
SUM=SUM+MN
DO 185 I=1,N
READ(9,87) (A(I,J), J=1,F)
185 READ(2,88) (S(I), (B(I,J), J=1,N))
READ(7,77) (BETA(I), I=1,N)
DO 187 I=1,N
187 READ(6,77) (D(I,J), J=1,N)
DO 190 I=1,N
190 SBETA=SBETA+S(I)*BETA(I)
PRINT *, NMAT, SBETA
DO 195 I=1,N
BETAB(I)=0.0
DO 195 K=1,N
195 BETAB(I)=BETAB(I)+BETA(K)*B(K,I)
GO 197 I=1,N
BETABA(I)=0.0
DO 197 K=1,N
197 BETARA(I)=BETARA(I)+BETAB(K)*A(K,I)
DO 200 I=1,N
200 BETABAU=BETABAU+BETABA(I)*UO(I)
PRINT *, NMAT, BETABAU
DO 205 I=1,N
205 BETBBET=BETBBET+BETAB(I)*BETA(I)

```

```

PRINT *,NMAT,9ETBET.
DO 207 I=1,P
SA(I)=C.O
DO 207 K=1,N
207 SA(I)=SA(I)+S(K)*A(K,I)
DO 209 I=1,P
209 SAU=SAU+SA(I)*UO(I)
PRINT *,NMAT,SAU

DO 210 I=1,N
DO 210 J=1,N
BD(I,J)=O.O
DO 210 K=1,N
210 BD(I,J)=BD(I,J)+B(I,K)*C(K,J)
DO 220 I=1,N
DO 220 J=1,N
220 SUMBD(I,J)=SUMBD(I,J)+BD(I,J)

DO 225 I=1,N
DO 225 J=1,R
BA(I,J)=O.O
DO 225 K=1,N
225 BA(I,J)=BA(I,J)+B(I,K)*A(K,J)
DO 227 I=1,R
DO 227 J=1,R
ABA(I,J)=O.O
DO 227 K=1,N
227 ABA(I,J)=ABA(I,J)+A(K,I)*BA(K,J)
DO 230 I=1,R
DO 230 J=1,R
ABAH(I,J)=C.O
DO 230 K=1,R
230 ABAH(I,J)=ABAH(I,J)+ABA(I,K)*H(K,J)

DO 233 I=1,R
DO 233 J=1,P
233 SUMABAH(I,J)=SUMABAH(I,J)+ABAH(I,J)
DO 235 I=1,R
DO 235 J=1,N
HA(I,J)=O.O
DO 235 K=1,P
235 HA(I,J)=HA(I,J)+H(I,K)*A(J,K)
DO 237 I=1,N
DO 237 J=1,N
BAHA(I,J)=O.O
DO 237 K=1,P
237 BAHA(I,J)=BAHA(I,J)+BA(I,K)*HA(K,J)

DO 240 I=1,N
DO 240 J=1,N
BAHABD(I,J)=O.O
DO 240 K=1,P
240 BAHABD(I,J)=BAHABD(I,J)+BAHA(I,K)*BD(K,J)
DO 242 I=1,N
DO 242 J=1,N
242 SUMB4D(I,J)=SUMB4D(I,J)+BAHABD(I,J)
DO 244 I=1,N
DO 244 J=1,N
DBAHABD(I,J)=O.O
DO 244 K=1,N
244 DBAHABD(I,J)=DBAHABD(I,J)+D(I,K)*BAHABD(K,J)
DO 245 I=1,N
DO 245 J=1,N
245 BETBET(I,J)=BETA(I)*BETA(J)
DO 247 I=1,N
DO 247 J=1,N
247 DELI(I,J)=DELI(I,J)+DBAHABD(I,J)+D(I,J)+BETBET(I,J)

DO 250 I=1,N
DO 250 J=1,N
B6D(I,J)=O.O
DO 250 K=1,N
250 B6D(I,J)=B6D(I,J)+BD(I,K)*BAHABD(K,J)
DO 260 I=1,N
DO 260 J=1,N
260 SUMB6D(I,J)=SUMB6D(I,J)+B6D(I,J)
180 CONTINUE

DO 265 I=1,N

```

```

DO 265 J=1,N
265 DEL1(I,J)=DEL1(I,J)/KK
DO 267 I=1,N
267 PRINT 77,(DEL1(I,J),J=1,N)

```

```

DO 272 I=1,N
272 PRINT 77,(SUM8D(I,J),J=1,N)
DO 273 I=1,N
273 PRINT 77,(SUMB4D(I,J),J=1,N)
DO 274 I=1,N
274 PRINT 77,(SUMB6D(I,J),J=1,N)
TP1=TP2=TP3=TR4=0.0

```

```

DO 278 I=1,R
278 TR1=TR1+SUMABA(I,I)
DO 280 I=1,N
TP2=TR2+SUMR4D(I,I)
TR3=TP3+SUMBD(I,I)
TR4=TP4+SUMB6D(I,I)
280 CONTINUE
DO 282 I=1,R
UABA(I)=0.0
DO 282 K=1,R
282 UABA(I)=UABA(I)+UO(K)*SUMABA(K,I)
DO 284 I=1,R
284 UABAU=UABAU+UABA(I)*UO(I)

```

```

PRINT 77,TR1,TR2,TR3,TR4
SIG(IJ)=0.0
SIG(IJ)=SIG(IJ)+SUMYS-2*(AU-2*BETA+UABAU+2*BETABAU+8*FIBBET
*+TR1*SICZ-TR2*SIGZ+TR3*SIGZ+TR4*SIGZ
SIG(IJ)=SIG(IJ)/SUM
PRINT *,SIG(IJ)

```

C--- CHECKING CONDITIONS FOR CONVERGENCE OF ITERATIVE PROCEDURE
C---

```

DA=(SIG(IJ)-SIGZ)/(SIG(IJ))
PRINT *,IJ,DA
IF (ABS(DA).GT.0.001) GO TO 281
281 DO 290 I=1,N
DO 290 J=1,N
DC(I,J)=(DEL1(I,J)-DEL(I,J))/(DEL1(I,J))
IF (ABS(DC(I,J)).GT.0.001) GO TO 99
290 CONTINUE
DO 300 I=1,N
PRINT 77,(DC(I,J),J=1,N)
300 WRITE(8,77) (DEL1(I,J),J=1,N)
PRINT 666,IJ
666 FORMAT(*ITERATION NUMBER:*,I3)
STOP
END

```