

The Maximum Likelihood Method for Non-Response in Sample Surveys

M.S. SRIVASTAVA and E.M. CARTER¹

ABSTRACT

The analysis of survey data becomes difficult in the presence of incomplete responses. By the use of the maximum likelihood method, estimators for the parameters of interest and test statistics can be generated. In this paper the maximum likelihood estimators are given for the case where the data is considered missing at random. A method for imputing the missing values is considered along with the problem of estimating the change points in the mean. Possible extensions of the results to structured covariances and to non-randomly incomplete data are also proposed.

KEY WORDS: Incomplete response; Missing at random; Maximum likelihood method; Imputation.

1. INTRODUCTION

Examples of non-response in sample surveys are in abundance. Various attempts with varying degrees of success have been made in the literature to solve this problem. The success of a particular procedure is dependent on the complexity of the problem. For example, when the data is not missing at random, the problem is far from being solved. The recent attempts by Heckman (1976) and Greenlees *et al.* (1982) among others, are highly sensitive to model misspecification. Similarly the hot-deck method has been severely criticized in the literature. However, when the sample size is large, the hot-deck method and a carefully designed regression method yield similar results in imputing the non-response income in Current Population Survey (CPS). See David, Little, Samuhel and Triest (1986).

The regression method is based on the assumption that the non-response is random, but unlike the hot-deck method does not require complete information from a previous census, which in a majority of cases is non-existent. Thus it appears that a carefully designed regression method may be of great help.

In this paper, the situation when the non-response is random is considered. Random non-response arises naturally in many situations. For example, in successive sampling, the sampling starts with a certain number of people from whom certain observations are obtained for a period of time. At the end of this period, some people are dropped from the survey and new people are added. The survey continues in this manner until completion. Examples of this nature are considered by Woolson, Leeper and Clarke (1978) and Woolson and Leeper (1980).

Even when the non-response is not random, the non-random nature of the incomplete data may be accounted for, by using a sufficient number of explanatory variables in the regression model and employing some of the techniques used in the hot-deck method as was done in David *et al.* (1986) for a univariate model. For example, in Section 2.5 a method for imputing the missing values is given.

¹ M.S. Srivastava, Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 1A1, and E.M. Carter, Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada N1G 2W1.

In the course of developing these results, a method will be derived for checking if there have been any changes over time in the response patterns. The models used can also be modified to include error variance-covariance matrices that are structured by the imposition of a time series to the response variables. In this paper it is assumed that the data are normally distributed from a simple random sampling scheme and that the data are missing at random. If the normality assumptions is dropped then the estimators can no longer be considered maximum likelihood estimators but may still be considered as good heuristic estimators.

In the next section, the form of the model will be described for the one sample problem.

2. THE ONE SAMPLE PROBLEM

2.1 The Model

The bivariate incomplete data problem is considered first to introduce the general procedure that follows. Let $\underline{y} = (y_1, y_2)'$ be a bivariate random vector with mean vector $\underline{\mu}$ and covariance matrix Σ . Without loss of generality, the missing data in the bivariate situation can be described as follows:

$$\begin{array}{l} y_{11}, \dots, y_{1n_1}, y_{1,n_1+1}, \dots, y_{1,n_1+n_2}, \text{ --- --- --- --- ---} \\ y_{21}, \dots, y_{2n_1}, \text{ --- --- --- --- ---} y_{2,n_1+n_2+1}, \dots, y_{2,n_1+n_2+n_3} \end{array} \quad (1)$$

That is, there are n_1 pairs of observations, n_2 observations on y_1 with the corresponding observation on y_2 missing, and n_3 observations on y_2 with the corresponding observation on y_1 missing. Thus $N = n_1 + n_2 + n_3$ observations are grouped into three subsets. If the complete data set were to be represented as $\underline{y}_1, \dots, \underline{y}_N$, then the actual observed responses can be defined as

$$\underline{z}_{1j} = B_1 \underline{y}_j = \underline{y}_j, \text{ for } j = 1, \dots, n_1,$$

$$\underline{z}_{2j} = B_2 \underline{y}_j = y_{1j}, \text{ for } j = n_1 + 1, \dots, n_1 + n_2,$$

and

$$\underline{z}_{3j} = B_3 \underline{y}_j = y_{2j}, \text{ for } j = n_1 + n_2 + 1, \dots, n_1 + n_2 + n_3,$$

where $B_1 = I_2$, the identity matrix, $B_2 = (1 \ 0)$ and $B_3 = (0 \ 1)$.

For the general multivariate one sample problem, there will be K subsets of the data containing n_1, \dots, n_K observations. Note that the maximum number of groups is $2^p - 1$. Also the total sample size is $N = n_1 + \dots + n_K$. If the k -th subset contains p_k characteristics i_1, \dots, i_{p_k} , then the matrix B_k would be a $p_k \times p$ matrix with a one in the (s, i_s) position for $s = 1, \dots, p_k$ and zero elsewhere. With this notation the observed vectors of responses can be written as:

$$\underline{z}_{kj} = B_k \underline{y}_{kj}, \quad j = 1, \dots, n_k, \quad k = 1, \dots, K.$$

Hence,

$$E(\underline{z}_{kj}) = B_k \underline{\mu},$$

and

$$\text{cov}(\underline{z}_{kj}) = B_k \Sigma B_k', \quad j = 1, \dots, n_k \text{ and } k = 1, \dots, K.$$

Example 1: (Data)

Wei and Lachin (1984) give the cholesterol levels for a treatment group studied at times 0, 6, 12, 20 and 24 months. For reasons not pertaining to the response variable, certain observations were incomplete. The data can be grouped into $K = 8$ subsets. For the first group of complete data the sample mean and covariance matrix, based on 36 observations, were:

$$\bar{\underline{z}}_1 = \begin{bmatrix} 226.6 \\ 249.6 \\ 252.6 \\ 253.1 \\ 256.7 \end{bmatrix}, \quad S_1 = \begin{bmatrix} 1964 & 1301 & 1151 & 960 & 1008 \\ 1301 & 1715 & 1109 & 1023 & 1199 \\ 1151 & 1109 & 1554 & 697 & 1266 \\ 960 & 1023 & 697 & 1148 & 667 \\ 1008 & 1199 & 1266 & 667 & 2546 \end{bmatrix}.$$

The data for each of the other subsets is given in Table 1 with the imputed values in parenthesis.

The matrices that define the model for the observed values are:

$$B_1 = I_5, \quad B_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad B_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$B_4 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad B_5 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad B_6 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix},$$

$$B_7 = (1 \ 0 \ 0 \ 0 \ 0),$$

$$B_8 = (0 \ 1 \ 0 \ 0 \ 0).$$

Now that the model is defined, estimation of the parameters and the imputation of the missing data can be performed.

2.2 Estimation of the Population Mean Vector and Covariance Matrix.

For each of the K subsets define the sample mean as

$$\bar{\underline{z}}_k = (n_k)^{-1} \sum_{j=1}^{n_k} \underline{z}_{kj}.$$

Table 1
Observed Cholesterol Levels and Imputed Values

		Variable	1	2	3	4	5
Subset 2:	$n_2 = 7$		224	273	242	274	(231)
			231	252	267	299	(233)
			268	296	314	330	(303)
			284	288	268	261	(300)
			217	231	276	257	(238)
			209	200	269	233	(323)
			200	261	264	300	(279)
Subset 3:	$n_3 = 1$		193	189	(257)	232	211
Subset 4:	$n_4 = 12$		201	219	220	(231)	(172)
			202	186	253	(245)	(328)
			209	207	167	(208)	(194)
			212	253	225	(157)	(194)
			276	326	304	(300)	(376)
			163	179	199	(211)	(224)
			239	243	265	(238)	(246)
			204	203	198	(234)	(171)
			247	211	225	(224)	(215)
			195	250	272	(265)	(231)
			228	228	279	(276)	(259)
			290	264	260	(249)	(325)
Subset 5:	$n_5 = 1$		227	247	(215)	(267)	220
Subset 6:	$n_6 = 5$		250	269	(327)	(250)	(295)
			175	214	(250)	(210)	(210)
			260	268	(327)	(248)	(321)
			197	218	(235)	(251)	(258)
			248	262	(286)	(251)	(271)
Subset 7:	$n_7 = 2$		193	(209)	(219)	(230)	(255)
			256	(277)	(294)	(260)	(281)
Subset 8:	$n_8 = 1$		(284)	327	(287)	(336)	(309)

Note: Total sample size is $N = 65$.

Then

$$E(\underline{\bar{z}}_k) = B_k \underline{\mu},$$

$$\text{cov}(\underline{\bar{z}}_k) = n_k^{-1} (B_k \Sigma B_k'),$$

and the $\underline{\bar{z}}_k$ are independently distributed for $k = 1, \dots, K$. Applying the least squares theory, we minimize

$$\sum_{k=1}^K \text{tr } n_k (B_k \Sigma B_k')^{-1} [\underline{\bar{z}}_k - B_k \underline{\mu}] [\underline{\bar{z}}_k - B_k \underline{\mu}]'.$$

The solution for a given value of Σ is

$$\hat{\underline{\mu}} = \left[\sum_{k=1}^K n_k B_k' (B_k \Sigma B_k')^{-1} B_k \right]^{-1} \left[\sum_{k=1}^K n_k B_k' (B_k \Sigma B_k')^{-1} \underline{\bar{z}}_k \right]. \quad (2)$$

If a normal distribution is assumed, then the least squares estimator is also the maximum likelihood estimator. Little (1982) has suggested the use of the EM algorithm for this problem and claimed that the normal distribution assumption is not necessary. That is, estimators of $\underline{\mu}$ and Σ can be defined as the solution of the normal likelihood equations even if the underlying population is not normal. These estimators cannot then be considered maximum likelihood estimators, but only heuristic estimators that are consistent under certain general conditions. However, if a normal distribution is not assumed, then there is no justification in maximizing the normal likelihood equations to obtain estimators. An alternative heuristic estimator for Σ is given at the end of this section. The maximum likelihood estimator for Σ , assuming normality, are given from Srivastava (1985) as the solution of the following equation:

$$H = \sum_{k=1}^K n_k B_k' (B_k \Sigma B_k')^{-1} B_k - \sum_{k=1}^K B_k' (B_k \Sigma B_k')^{-1} V_k (B_k \Sigma B_k')^{-1} B_k = 0, \quad (3)$$

where

$$V_k = (\underline{z}_{k1} - B_k \underline{\mu}, \dots, \underline{z}_{k,n_k} - B_k \underline{\mu}) (\underline{z}_{k1} - B_k \underline{\mu}, \dots, \underline{z}_{k,n_k} - B_k \underline{\mu})'.$$

Methods for computing the solutions of (2) and (3) are given in Section 3.

Note: Alternate estimators for the covariance matrix can be defined heuristically without the normality assumption. For example $\hat{\Sigma}$ can be defined as the value of Σ that minimizes

$$\sum_{k=1}^K n_k^{-1} \text{tr}[(B_k \Sigma B_k')^{-1} V_k - n_k I_k]^2 \quad (4)$$

However, the covariance matrix must be positive definite; therefore any expression that is minimized must yield a positive definite solution. If one of the groups contains complete data, then (4) will be infinite for any singular matrix Σ ; hence, there will exist a minimum for (4) in the space of positive definite matrices. A similar argument holds for the maximum likelihood estimators.

2.3 Asymptotic Distribution of $\hat{\underline{\mu}}$.

From (2) it follows that $\hat{\underline{\mu}}$ is asymptotically normally distributed with mean $\underline{\mu}$ and covariance matrix

$$P = \left[\sum_{k=1}^K n_k B_k' (B_k \Sigma B_k')^{-1} B_k \right]^{-1}, \quad (5)$$

which can be estimated by \hat{P} obtained from P by substituting the $\hat{\Sigma}$ for Σ . Using this asymptotic theory, tests of significance and confidence regions (intervals) for $\underline{\mu}$ or linear combinations of $\underline{\mu}$ can be obtained. Alternatively, the likelihood ratio tests given by Srivastava (1985) may be used for testing the hypothesis $H: \underline{\mu} = \underline{0}$ against the alternative $A: \underline{\mu} \neq \underline{0}$. The likelihood ratio test rejects the null hypothesis H if

$$\lambda = \prod [|B_k \hat{\Sigma} B_k'| / |B_k \tilde{\Sigma} B_k'|]^{n_k/2} > \chi_{p, \alpha}^2,$$

where $\tilde{\Sigma}$ is the MLE of Σ under H and $\chi_{p, \alpha}^2$ is the upper $100\alpha\%$ point of a chi-square distribution with p degrees of freedom.

2.4 Maximum Likelihood Estimates for Example 1

The maximum likelihood estimates for example 1 were obtained as:

$$\hat{\underline{\mu}} = \begin{bmatrix} 226.82 \\ 246.78 \\ 252.02 \\ 255.15 \\ 255.22 \end{bmatrix} \quad \text{and} \quad \hat{\Sigma} = \begin{bmatrix} 1809 & 1220 & 1033 & 873 & 913 \\ 1220 & 1642 & 992 & 1017 & 1121 \\ 1033 & 992 & 1438 & 718 & 1189 \\ 873 & 1017 & 718 & 1233 & 915 \\ 913 & 1121 & 1189 & 915 & 2508 \end{bmatrix}.$$

The estimated covariance matrix for the estimate of the mean vector is

$$P^{-1} = \begin{bmatrix} 28.05 & 18.78 & 15.96 & 13.46 & 14.08 \\ 18.78 & 25.67 & 15.42 & 15.84 & 17.51 \\ 15.96 & 15.42 & 24.19 & 11.24 & 19.31 \\ 13.46 & 15.84 & 11.24 & 23.33 & 15.38 \\ 14.08 & 17.51 & 19.31 & 15.38 & 54.77 \end{bmatrix}.$$

Inference on $\underline{\mu}$ can be made from the asymptotic distribution of the estimators given in Section 2.3.

2.5 Imputation

The imputation of the missing data can be made from the conditional distribution of the unobserved data given the observed data. That is define the matrices C_k for $k = 1, \dots, K$

to be the complements of B_k . That is for a $p_k \times p$ matrix B_k with ones as the (s, i_s) entries for $s = 1, \dots, p_k$ and 0's elsewhere, the matrix C_k is defined as the $(p - p_k) \times p$ matrix with ones in the (t, i_t) position and 0's elsewhere for $i_t \neq i_s$ for all $t = 1, \dots, (p - p_k)$ and $s = 1, \dots, p_k$. If the response vector \underline{y}_{kj} corresponds to the j -th observation from subset k , then the actual observed response vector is $\underline{z}_{kj} = B_k \underline{y}_{kj}$ and the unobserved vector is $\underline{\hat{u}}_{kj} = C_k \underline{y}_{kj}$. The estimated value for the missing vector is given by

$$\underline{\hat{u}}_{kj} = C_k \underline{\hat{\mu}} + [C_k \hat{\Sigma} B_k'] [B_k \hat{\Sigma} B_k']^{-1} (\underline{z}_{kj} - B_k \underline{\hat{\mu}}) \quad (6)$$

Note that the estimated values for the missing vector have no random error. If the data is to be used at a subsequent analysis, with these imputed values, as if it were a complete data set, then the estimated error covariance matrix will be too small. The problem of underestimating the covariance matrix can be overcome by adding in an appropriate residual $\underline{\epsilon}$ to the estimated value $\underline{\mu}_{kj}$. If the first subset of complete data is sufficiently large then the residual vectors for missing observations in subset k can be randomly drawn from the set of values

$$(C_k \underline{y}_{li} - C_k \underline{\hat{\mu}}) - [C_k \hat{\Sigma} B_k'] [B_k \hat{\Sigma} B_k']^{-1} (B_k \underline{y}_{li} - B_k \underline{\hat{\mu}}) \text{ for } i = 1, \dots, n_1. \quad (7)$$

Example 1 (continued):

The complete data set, including the imputed values based on (6) and (7) are given in Table 1 for subsets 2-8 with the imputed values in parenthesis.

3. COMPUTATIONAL PROCEDURES

Equations (2) and (3) can be solved iteratively. A procedure using a combined Newton-Raphson and steepest ascent method is given in Carter (1986) for a general case that includes linearly restricted means and covariances. The procedure is a generalization of the one given by Hartley and Hocking (1971). The method can be described as follows. For an initial choice of Σ , say Σ_0 , suppose

$$\Sigma = \Sigma_0 + \Lambda$$

is a solution. This expression is substituted into (3) and the equation is then expanded in a series involving only the linear terms of Λ . The following approximate solution for Λ results. Define

$$Q = \sum_{k=1}^K (D_k \otimes D_k - D_k \otimes F_k - F_k \otimes D_k),$$

where $A \otimes B$ denotes the kronecker product of two matrices A and B defined by $A \otimes B = (a_{ij} B)$,

$$D_k = B_k' (B_k \Sigma_0 B_k')^{-1} B_k,$$

and

$$F_k = B_k' (B_k \Sigma_0 B_k')^{-1} V_k (B_k \Sigma_0 B_k')^{-1} B_k.$$

For any matrix $A = (\underline{a}_1, \dots, \underline{a}_q)'$, we define $\text{vec}(A) = (\underline{a}_1', \dots, \underline{a}_q')'$. Then (3) can be written as approximately

$$Q \text{vec}(\Lambda) = \text{vec}(E),$$

where

$$E = \sum_{k=1}^K (D_k - F_k).$$

To insure the nonsingularity of Q , we shall write the solution for $\text{vec}(\Lambda)$ as

$$\text{vec}(\Lambda) = (Q + \lambda I)^{-1} \text{vec}(E), \quad (8)$$

where λ is allowed to vary with the algorithm but is initially set to a very small number. For a given value of Σ , $\hat{\mu}$ is obtained from (2) and then a value of Λ is obtained from (8) to produce an updated estimate for Σ . The procedure is then iterated until a desired level of convergence is reached.

The above method can be extended to more complex structured covariance matrices; however, the procedure does require the inversion of $Q + \lambda I$. For a large number of variables this matrix will be extremely large. In this instance the alternate method of solving (3) using the EM algorithm is preferable. Again the procedure is iterative, so calculations must be performed using the updated estimates of μ and Σ at each iteration. For an initial choice of Σ say Σ_0 , define the complete predicted vector $\hat{y}_{kj} = B_k' \underline{z}_{kj} + C_k' \hat{\mu}_{kj}$, where the predicted missing value $\hat{\mu}_{kj}$ is given in (6). Then

$$\hat{\mu} = (1/N) \sum_{k=1}^K \sum_{j=1}^n \hat{y}_{kj}$$

Define the matrix V by

$$V = \sum_{k=1}^K \sum_{j=1}^{n_k} (\hat{y}_{kj} - \hat{\mu}) (\hat{y}_{kj} - \hat{\mu})'.$$

The updated estimate of Σ is then given by

$$\hat{\Sigma} = (1/N) [V + \sum_{k=1}^K n_k C_k' H_k C_k],$$

where H_k is the conditional variance of the incomplete data given the observed data for the k -th class defined by

$$H_k = C_k \Sigma C_k' - (C_k \Sigma B_k') (B_k \Sigma B_k')^{-1} (B_k \Sigma C_k').$$

The procedure is then iterated. The EM algorithm is advantageous for those situations where there exists simple closed form solutions for the likelihood equations in the complete data situations. If a Newton-Raphson procedure is necessary to solve the complete data likelihood equations then little is gained from the EM algorithm.

4. A REGRESSION MODEL

4.1 Incomplete Response Variables.

The model discussed in section 2 can be extended to handle the regression situation. The data is again partitioned into K subsets. Then the following regression model is formed:

$$Z_k = B_k' \beta A_k + \epsilon_k, \text{ for } k = 1, \dots, K,$$

where Z_k is a $p_k \times n_k$ matrix of observed values, β is a $p \times q$ matrix of unknown parameters, B_k is as defined in Section 2, A_k is the design matrix for the matrix Z_k and the columns of ϵ_k are independently distributed with mean $\underline{0}$ and covariance matrix $B_k \Sigma B_k'$. For a given Σ , the least squares estimator of β can be written from Carter (1986) explicitly as

$$\text{vec } \hat{\beta} = P^{-1} \text{vec}(E),$$

where

$$P = \sum_{k=1}^K n_k B_k' (B_k \Sigma B_k')^{-1} B_k \otimes A_k A_k', \quad (10)$$

$$E = \sum_{k=1}^K B_k' (B_k \Sigma B_k')^{-1} Z_k A_k'. \quad (11)$$

The maximum likelihood estimator of Σ is given by the same formula as (3), except that now

$$V_k = [Z_k - B_k \beta A_k] [Z_k - B_k \beta A_k]'. \quad (12)$$

The asymptotic distribution of $\hat{\beta}$ can be written in the form

$$\text{vec}(\hat{\beta}) \sim N_{pq}(\text{vec}(\beta), P^{-1}). \quad (13)$$

Inference on the regression parameters can be made from this asymptotic distribution or from the likelihood ratio statistic given in Srivastava (1985).

4.2 Incomplete Explanatory Variables

In Section 3.1, the design matrices were assumed to be known completely. In some instances the explanatory variables can also be incomplete. If the explanatory variables are random, then these missing values can first be imputed for the explanatory variables given the observed data, using the procedure of Section 2. Once imputed values for the explanatory variables are obtained then the method of Section 3.1 can be applied to estimate the regression parameters and to impute the missing response variables.

4.3 The Likelihood Ratio Test.

The likelihood ratio procedure can be used to determine if the variables in the model are significant. To test the hypothesis

$$H: \beta = \beta_1 F \text{ vs } A: \beta \neq \beta_1 F,$$

for F an $m \times q$ matrix of full rank, the estimates of Σ are obtained under the null hypothesis ($\tilde{\Sigma}$) and under the alternate hypothesis ($\hat{\Sigma}$). The null hypothesis is rejected at the α level of significance if

$$-2 \ln \lambda > \chi^2_{(q-m)p; \alpha},$$

where

$$\lambda = \prod_{k=1}^K |B_k \hat{\Sigma} B_k'|^{n_k/2} / |B_k \tilde{\Sigma} B_k'|^{n_k/2}. \quad (14)$$

5. ESTIMATING A CHANGE POINT

Consider a sequence of observations y_j , $j = 1, \dots, N$, with expected values $E(y_j) = \mu_j$. Srivastava and Worsley (1986) have given a procedure for estimating the point of change of the mean vectors μ_j . It is first assumed that the change occurs at some point r . Then the following hypothesis is tested.

$$H: \mu_1 = \dots = \mu_N$$

$$A: \mu_1 = \dots = \mu_r \neq \mu_{r+1} = \dots = \mu_N.$$

The likelihood ratio statistic is then calculated as λ_r , for $r = 1, \dots, N-1$. The estimated point of change is that value of r that yields the maximum value of λ_r .

The existence of incomplete data poses no problems for estimating the change point. The linear model is set up as for the complete data case, then the observations are grouped into the K subsets. Suppose that the observed portion of y_j is z_{ki} . Then under the alternate hypothesis for a given r , $\hat{\Sigma}$ the estimate for Σ is given from (3) for the regression model defined in (9)–(12), where the parameter matrix β is defined as

$$\beta = (\mu_1, \mu_2)$$

and the design matrix for the k -th subset is defined by

$$A_k = \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 \end{bmatrix},$$

where the i -th column of A_k has a one in the first row if observation z_{ki} corresponds to the vector y_j and $j \leq r$ and zero otherwise. Under the null hypothesis the population mean vector is considered the same for all N observations; hence, Σ the estimate for Σ is given from (2) and (3) for the one population mean problem. The likelihood ratio statistic is obtained from (14).

Modifications of this procedure are possible. For example the vectors y_j for $j = 1, \dots, N$ could be sample means for N sampling time points. Multiple change points can be obtained by repeating the procedure on each section of the data. For 50 observations, if the change point occurs at point 20 then the procedure is repeated for points 1-20 and 21-50.

6. STRUCTURED COVARIANCE MATRICES

For longitudinal studies the error vectors over time may not be arbitrary, but may follow a time series model. If such a model can be assumed, then the number of parameters to be estimated is reduced. A stationary time series would assume that the covariance matrix Σ can be written as

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 \dots & \rho_{p-2} \\ . & & & . \\ . & & & . \\ . & & & . \\ \rho_{p-1} & \rho_{p-2} & \dots & \rho_1 & 1 \end{bmatrix}. \quad (15)$$

Further models can be obtained. The correlations ρ_j can be structured. For example ρ_j can be set equal to $\rho^{|j|}$. The likelihood equations can be solved using the Newton-Raphson technique. Carter (1986) considered the case where the covariance matrix can be written as $\text{vec}(\Sigma) = G\gamma$ for some matrix G . By defining $\gamma_i = \sigma^2 \rho_i$ for $i = 1, \dots, p-1$ and $\gamma_p = \sigma^2$, then the covariance matrix for the stationary time series can be expressed in this linearly restricted form. For example for $p = 3$ we have

$$\begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{13} \\ \sigma_{21} \\ \sigma_{22} \\ \sigma_{23} \\ \sigma_{31} \\ \sigma_{32} \\ \sigma_{33} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix}$$

The estimate of Σ can be solved numerically from the likelihood equation $G'H = 0$, where H is defined in (3). Numerically the Newton-Raphson algorithm from Section 3 can be employed with the modification that the estimate for $\underline{\gamma}$ at each iteration is given by

$$\hat{\underline{\gamma}} = (G'QG + \lambda I)^{-1} G' \text{vec}(E).$$

REFERENCES

- CARTER, E.M. (1986). The analysis of a generalized multivariate linear model. Technical Report, University of Guelph.
- DAVID, M., LITTLE, R.J., SAMUHEL, M.E., and TRIEST, R.K. (1986). Alternative methods for CPS income imputation. *Journal of the American Statistical Association*, 81, 29-41.
- DEMPSTER, A.P., LAIRD, N.M., and RUBIN, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Ser. B*, 39, 1-38.
- DRAPER and SMITH (1981). *Applied Regression Analysis*. New York: Wiley.
- GREENLEES, W.S., REECE, J.S., and ZIESCHANG, K.D. (1982). Imputation of missing values when the probability of response depends on the variable being imputed. *Journal of the American Statistical Association*, 77, 251-261.
- HARTLEY, H.O., and HOCKING, R.R. (1971). The analysis of incomplete data (with discussion). *Biometrics*, 27, 783-823.
- HECKMAN, J.D. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimation for such models. *Annals of Economic and Social Measurements*, 5, 475-492.
- LITTLE, R.J.A. (1982). Models for non-response in sample surveys. *Journal of the American Statistical Association*, 77, 237-250.
- RUBIN, D.B., and SZATROWSKI, T.H. (1982). Finding maximum likelihood estimates of patterned covariance matrices by the EM algorithm. *Biometrika*, 69, 657-660.
- SRIVASTAVA, M.S. (1985). Multivariate data with missing observations. *Communications in Statistics Theory and Methods*, 14, 775-792.
- SRIVASTAVA, M.S., and WORSLEY, K.J. (1986). Likelihood ratio tests for a change in the multivariate mean. *Journal of the American Statistical Association*, 81, 199-204.
- WEI, L.J., and LACHIN, J.M. (1984). Two sample asymptotically distribution free tests for incomplete multivariate observations. *Journal of the American Statistical Association*, 79, 653-661.
- WOOLSON, R.F., and LEEPER, J.D. (1980). Growth curve analysis of complete and longitudinal data, *Communications in Statistics, Theory and Methods*, 9, 1491-1513.
- WOOLSON, R.F., LEEPER, J.D., and CLARKE, W.R. (1978). Analysis of incomplete data from longitudinal and mixed longitudinal studies. *Journal of the Royal Statistical Society, Ser. A*, 141, 242-252.