Sankhyā : The Indian Journal of Statistics
1982, Volume 44, Series B, Pt. 2 pp. 143–153.

ESTIMATION OF COVARIANCE FROM UNBALANCED DATA

By A. K. GUPTA and V. K. ROHATGI
Bowling Green State University

SUMMARY. In this paper the problem of estimation of the covariance of a bivariate normal distribution when data is unbalanced has been considered. Besides the naive and the maximum likelihood estimators, three regression estimators have been proposed and their properties investigated. Comparative study, in terms of their variances, has been carried out to establish the superiority of the regression-estimators for the given number of complete and incomplete observations.

1. Introduction

Let the random variables X and Y be jointly distributed with $E(X) = \mu_1$, $E(Y) = \mu_2$, $V(X) = \sigma_1^2$, $V(Y) = \sigma_2^2$ and $\text{cov}(X, Y) = \gamma$. If ρ is the correlation coefficient between X and Y, then $\gamma = \rho\sigma_1\sigma_2$. Suppose (x_i, y_i), $i = 1, 2, \ldots, n$ are n pairs of observations on (X, Y) and x_{n+j}, $j = 1, \ldots, n_1$ are n_1 additional observations on X only, that is, the observations on Y are missing or simply cannot be obtained. Although the missing observations are assumed to occur at random one can, without loss of generality, arrange them in the above order. In this paper we consider the problem of estimating γ when the data is unbalanced.

In the next section we study a naive estimator, the maximum likelihood estimator and three regression-estimators based on predicted values of Y's. In Section 3 a comparative study, in terms of relative efficiency, of these estimators has been carried out. The case when n_2 single observations on Y are also available has also been studied in Section 4.

2. Estimation of γ

Let us write

$$\bar{X}^{(n)} = n^{-1} \sum_{i=1}^{n} X_i, \quad \bar{X}^{(n_1)} = n_1^{-1} \sum_{i=1}^{n_1} X_{n+i}, \quad \bar{Y}^{(n)} = n^{-1} \sum_{i=1}^{n} Y_i,$$

$$\bar{X}^{(n+n_1)} = (n+n_1)^{-1}[n\bar{X}^{(n)} + n_1\bar{X}^{(n_1)}],$$

$$S_{11} = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}^{(n)})^2,$$

$$S_{22} = (n-1)^{-1} \sum_{i=1}^{n} (Y_i - \bar{Y}^{(n)})^2,$$

and

$$S_{12} = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}^{(n)})(Y_i - \bar{Y}^{(n)}). \qquad \ldots \quad (1)$$

In this section we propose five estimators of γ, four of which are distribution-free. However, for comparing their efficiencies we will assume (in Section 3) that the sample comes from a bivariate normal population.

(i) The first estimator is a naive estimator, which is also a special case of the estimator proposed by Boas (1967) ($n_2 - n = 0$, in his notation) and is given by

$$\hat{\gamma}_{B_0} = S_{12}, \quad n > 1. \qquad \ldots \quad (2)$$

It is noted that the estimator $\hat{\gamma}_{B_0}$ is unbiased and consistent with variance given by

$$V(\hat{\gamma}_{B_0}) = n^{-1}\left[\mu_{22} - \frac{n-2}{n-1}\gamma^2 + \frac{\sigma_1^2\sigma_2^2}{n-1}\right], \qquad \ldots \quad (3)$$

where $\mu_{22} = E[(X-\mu_1)^2(Y-\mu_2)^2]$.

(ii) Under the assumption of normality, Anderson (1957) (see also Olkin and Sylvan, 1977) has derived the maximum lieklihood estimators of the parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$ and ρ. From his results the maximum likelihood estimator of γ can be easily shown to be

$$\hat{\gamma}_A = \frac{1}{n+n_1} \frac{S_{12}}{S_{11}} \sum_{i=1}^{n+n_1} (X_i - \bar{X}^{(n+n_1)})^2. \qquad \ldots \quad (4)$$

This estimator is biased. An unbiased estimator based on $\hat{\gamma}_A$ is given by

$$\hat{\gamma} = \frac{1}{n+n_1-1} \frac{S_{12}}{S_{11}} \sum_{1}^{n+n_1} (X_i - \bar{X}^{(n+n_1)})^2. \qquad \ldots \quad (5)$$

The variance of $\hat{\gamma}$ can be easily obtained as

$$V(\hat{\gamma}) = \frac{\sigma_1^2\sigma_2^2}{(n+n_1-1)^2} [(n-1)(1+\rho^2 n) + 2n_1(1+n\rho^2-2\rho^2)$$

$$+ n_1(n_1+2)\Delta(\rho)] - \rho^2\sigma_1^2\sigma_2^2. \qquad \ldots \quad (6)$$

where $\Delta(\rho) = (1+n\rho^2-4\rho^2)/(n-3)$, $n > 3$

(iii) The next three estimators are based on the prediction of Y-values from the estimated regression line. The precision of an estimator of the population mean of X (say), can, in general, be increased by the use of an auxiliary variable Y which is correlated with X. If the relationship is a straight line, a linear regression-estimator can be constructed. In many cases when the relationship between X and Y is nonlinear, it is possible to transform X and Y into two new variables U and V, which are linearly related. In the present context we will use the regression technique to predict the n_1 missing values of Y given the values of X. This method was suggested by the authors in Gupta and Rohatgi (1977, 1978).

According to Gupta and Rohatgi (1977, 1978) we use the regression line

$$\hat{Y}_i = \bar{Y}^{(n)} + \frac{S_{12}}{S_{11}} (X_i - \bar{X}^{(n+n_1)}),\ i = n+1, \ldots, n+n_1, \quad \ldots (7)$$

for predicting the missing Y-values. Then the first regression-estimator we suggest is a pooled estimator given by

$$\hat{\gamma}_1 = \Big[\sum_1^n (X_i - \bar{X}^{(n)})(Y_i - \bar{Y}^{(n)}) + \sum_{n+1}^{n+n_1} (X_i - \bar{X}^{(n_1)})(\hat{Y}_i - \hat{\bar{Y}}^{(n_1)})\Big] / (n+n_1-2) \quad \ldots (8)$$

where

$$\hat{\bar{Y}}^{(n_1)} = \bar{Y}^{(n)} + \frac{S_{12}}{S_{11}} (\bar{X}^{(n_1)} - \bar{X}^{(n+n_1)}). \quad \ldots (9)$$

It may be noted that $\hat{\gamma}_1$ remains unchanged if we estimate μ_1 in (7) by $\bar{X}^{(n)}$ in place of $\bar{X}^{(n+n_1)}$. The expression for $\hat{\gamma}_1$ in (8) can be simplified to

$$\hat{\gamma}_1 = S_{12} \Big[n - 1 + \sum_{n+1}^{(n+n_1)} (X_i - \bar{X}^{(n_1)})^2 / S_{11}\Big] / (n+n_1-2). \quad \ldots (10)$$

The second regression-estimator we propose is also based on the estimated regression line (7) and is given by

$$\hat{\gamma}_2 = k_2 \Big[\sum_1^n (X_i - \bar{X}^{(n+n_1)}) Y_i + \sum_{n+1}^{n+n_1} (X_i - \bar{X}^{(n+n_1)}) \hat{Y}_i\Big] \quad \ldots (11)$$

where $k_2 = \left[n+n_1-2+\frac{n}{n+n_1}\right]^{-1}$. This estimator uses the improved estimator $\bar{X}^{(n+n_1)}$ for μ_1 and can be simplified to

$$\hat{\gamma}_2 = k_2\left[(n-1)S_{12}+\frac{S_{12}}{S_{11}}\left\{\sum_{n+1}^{n+n_1}(X_i-\bar{X}^{(n_1)})^2+\frac{n^2 n_1}{(n+n_1)^2}(\bar{X}^{(n_1)}-\bar{X}^{(n)})^2\right\}\right]. \quad \dots \quad (12)$$

However, if one uses $\bar{X}^{(n)}$ instead of $\bar{X}^{(n+n_1)}$ in (7) to estimate μ_1 i.e.,

$$\hat{Y}_i = \bar{Y}^{(n)}+\frac{S_{12}}{S_{11}}(X_i-\bar{X}^{(n)}) \quad \dots \quad (13)$$

then $\hat{\gamma}_2 \equiv \hat{\gamma}$, the estimator based on maximum likelihood estimation given earlier by (5). The estimator $\hat{\gamma}_2$ can be written in terms of $\hat{\gamma}$ as follows

$$\hat{\gamma}_2 = k_2\left[(n+n_1-1)\hat{\gamma}-\frac{nn_1^2}{(n+n_1)^2}\frac{S_{12}}{S_{11}}(\bar{X}^{(n)}-\bar{X}^{(n_1)})^2\right]. \quad \dots \quad (14)$$

The third regression-estimator is also based on (7) and is given by

$$\hat{\gamma}_3 = k_3\sum_1^n(X_i-\bar{X}^{(n+n_1)})(Y_i-\hat{\bar{Y}}^{(n+n_1)}) \quad \dots \quad (15)$$

where

$$k_3 = \left[n-1+\frac{nn_1}{(n+n_1)^2}\right]^{-1} \text{ and } \hat{\bar{Y}}^{(n+n_1)} = (n\bar{Y}^{(n)}+n_1\hat{\bar{Y}}^{(n_1)})/(n+n_1).$$

$\hat{\gamma}_3$ can be simplified to obtain

$$\hat{\gamma}_3 = k_3\left[(n-1)S_{12}+\frac{n^2n_1^2}{(n+n_1)^3}\frac{S_{12}}{S_{11}}(\bar{X}^{(n)}-\bar{X}^{(n_1)})^2\right]. \quad \dots \quad (16)$$

The four estimators $\hat{\gamma}_{B_0}$, $\hat{\gamma}_i$ ($i = 1, 2, 3$) are distribution free whereas the estimator based on the maximum likelihood estimator, $\hat{\gamma}$, assumes normality. In the next section we study their properties and relative efficiencies in the special case when (X, Y) has a bivariate normal distribution.

3. THE BIVARIATE NORMAL CASE

To study the properties of and to compare the estimators $\hat{\gamma}_{B_0}$, $\hat{\gamma}$, $\hat{\gamma}_i$ $(i = 1, 2, 3)$ we assume that (X, Y) has a bivariate normal distribution. We first note that $(\bar{X}^{(n_1)}, \bar{Y}^{(n_1)}, \bar{X}^{(n_1)}, S_{11}, S_{22}, S_{12}, S'_{22})$ where

$$S'_{22} = (n_1-1)^{-1} \sum_{i=n+1}^{n+n_1} (X_i - \bar{X}^{(n_1)})^2$$

is minimal sufficient. All the five estimators being considered here are functions of the minimal sufficient statistic.

In view of (3) we have

$$V(\hat{\gamma}_{B_0}) = \sigma_1^2\sigma_2^2(1+\rho^2)/(n-1), \qquad \dots \quad (17)$$

and the variance of $\hat{\gamma}$ is given by (6). The unbiasedness of $\hat{\gamma}_i$ $(i = 1, 2, 3)$ follows from the fact that $E(S_{11}^{-1}S_{12} \mid X_1, \dots, X_{n+n_1}) = \gamma\sigma_1^{-2}$. The expressions for the variances of $\hat{\gamma}_i$ $(i = 1, 2, 3)$ are easy to derive. Indeed, for $n > 3$, we have

$$V(\hat{\gamma}_1) = \frac{\sigma_1^2\sigma_2^2}{(n+n_1-2)^2} [(n-1)(1+\rho^2 n)+(n_1^2-1)\Delta(\rho)$$
$$+2(n_1-1)(1+n\rho^2-2\rho^2)]-\rho^2\sigma_1^2\sigma_2^2 \qquad \dots \quad (18)$$

$$V(\hat{\gamma}_2) = k_2^2\sigma_1^2\sigma_2^2\Big[(n-1)(1+n\rho^2)+\Delta(\rho)$$
$$\cdot\Big\{(n_1^2-1)+\frac{2n(n_1-1)}{n+n_1}+\frac{3n^2}{(n+n_1)^2}\Big\}+2(1+n\rho^2-2\rho^2)$$
$$\cdot\Big\{n_1-1+\frac{n}{n+n_1}\Big\}\Big]-\rho^2\sigma_1^2\sigma_2^2 \qquad \dots \quad (19)$$

and

$$V(\hat{\gamma}_3) = k_3^2\sigma_1^2\sigma_2^2\Big[(n-1)(1+n\rho^2)+\Delta(\rho)\ \frac{3n^2n_1^2}{(n+n_1)^4}+\frac{2nn_1}{(n+n_1)^3}$$
$$\cdot(1+n\rho^2-2\rho^2)\Big]-\rho^2\sigma_1^2\sigma_2^2. \qquad \dots \quad (20)$$

Let us now define the exact relative efficiency of an unbiased estimator h with respect to another unbiased estimator g of γ, both with finite variances, by

$$e(h : g) = \frac{V(g)}{V(h)}. \quad \dots \quad (21)$$

Then the estimator h is more efficient than g if $e(g : h) < 1$. From (6), and (16) through (19) we see after simple computations that for $n > 3$,

$$e(\hat{\gamma}_{B_0} : \hat{\gamma}) < 1 \Longleftrightarrow \rho^2 > 1/(n-2), \quad \dots \quad (22)$$

$$e(\hat{\gamma}_{B_0} : \hat{\gamma}_1) < 1 \Longleftrightarrow \rho^2 > 1/(n-2), \quad \dots \quad (23)$$

and

$$e(\hat{\gamma} : \hat{\gamma}_1) < 1 \Longleftrightarrow \rho^2 < 1/(n-2). \quad \dots \quad (24)$$

It follows that we need to compare the performance of $\hat{\gamma}_{B_0}$ vs. $\hat{\gamma}_2$ and $\hat{\gamma}_3$ for $\rho^2 < 1/(n-2)$ and that of $\hat{\gamma}$ vs. $\hat{\gamma}_2$ and $\hat{\gamma}_3$ for $\rho^2 > 1/(n-2)$.

Moreover, letting n and $n_1 \to \infty$ such that $n_1/(n+n_1) \to \lambda_1$ we see that

$$V(\hat{\gamma}_2)/V(\hat{\gamma}) \to 1$$

whereas

$$V(\hat{\gamma}_3)/V(\hat{\gamma}) \to (1+c)^2$$

where $c = \lambda_1/(1-\lambda_1)$. This is to be expected since $\hat{\gamma}$, being essentially the maximum likelihood estimate, is asymptotically efficient. Therefore we need only compare the estimators for small and moderate values of n. Accordingly we computed the relative efficiencies $e(\hat{\gamma} : \hat{\gamma}_i)$, $i = 1, 2, 3$ for $n = 8, 10, 15, 20$; $\rho^2 = 0, 0{\cdot}1(0{\cdot}2)\,0{\cdot}9$; and $n_1/(n+n_1) = \lambda_1 = 0{\cdot}1(0{\cdot}2)\,0{\cdot}9$. These results are presented in Table 1 for $n = 8$ only.

TABLE 1. RELATIVE EFFICIENCY OF $\hat{\gamma}_1$, $\hat{\gamma}_2$ AND $\hat{\gamma}_3$ WITH RESPECT TO $\hat{\gamma}$ FOR $n = 8$

λ_1	ρ^2	$e(\hat{\gamma} : \hat{\gamma}_1)$	$e(\hat{\gamma} : \hat{\gamma}_2)$	$e(\hat{\gamma} : \hat{\gamma}_3)$
0.1	0.0	.9507	.9917	.9574
	0.1	.9816	.9932	.9800
	0.3	1.0326	.9956	1.0174
	0.5	1.0731	.9976	1.0470
	0.7	1.1060	.9992	1.0711
	0.9	1.1332	1.0005	1.0911
0.3	0.0	.9748	.9879	.8862
	0.1	.9901	.9903	.9475
	0.3	1.0191	.9948	1.0638
	0.5	1.0460	.9989	1.1721
	0.7	1.0712	1.0029	1.2733
	0.9	1.0948	1.0065	1.3680
0.5	0.0	.9890	.9920	.8273
	0.1	.9955	.9940	.9209
	0.3	1.0094	.9981	1.1219
	0.5	1.0248	1.0027	1.3439
	0.7	1.0419	1.0078	1.5903
	0.9	1.0610	1.0135	1.8655
0.7	0.0	.9966	.9969	.7767
	0.1	.9985	.9979	.8979
	0.3	1.0033	1.0003	1.1909
	0.5	1.0096	1.0036	1.5791
	0.7	1.0184	1.0081	2.1178
	0.9	1.0314	1.0147	2.9155
0.9	0.0	1.0000	1.0000	.7332
	0.1	.9998	.9998	.8794
	0.3	1.0003	1.0002	1.2752
	0.5	1.0012	1.0010	1.9149
	0.7	1.0027	1.0021	3.1240
	0.9	1.0067	1.0054	6.2738

We note that for $\rho^2 \leqslant 1/(n-2)$ (same result holds for $n = 10, 15, 20$) $\hat{\gamma}_3$ is preferable. It therefore suffices to compare $\hat{\gamma}_{B_0}$ and $\hat{\gamma}_3$ for $\rho^2 \leqslant 1/(n-2)$. This was done for $n = 8, 10, 15, 20$ and same set of parameter values of ρ^2 and λ_1 as used above. Table 2 gives a summary of our recommendations.

TABLE 2. RECOMMENDATIONS

ρ^2	preferred estimator
0	$\hat{\gamma}_{B_0}$
$0 < \rho^2 < 1/(n-2)$	$\hat{\gamma}_s$
$\rho^2 > 1/(n-2)$	
0.3	$\hat{\gamma}_2$ for $n = 8, 10, \lambda_1 = .1, .3, .5$; $n = 15, \lambda_1 = .1, .3$, and $n = 20, \lambda_1 = .1$; $\hat{\gamma}$ otherwise
0.5	$\hat{\gamma}_2$ for $n = 8, 10, 15, 20, \lambda_1 = .1$ and $n = 8, \lambda_1 = .3$; $\hat{\gamma}$ otherwise
0.7	$\hat{\gamma}_2$ for $n = 8, 10, \lambda_1 = .1$; $\hat{\gamma}$ otherwise
0.9	$\hat{\gamma}$ for all n and λ_1

4. The $n_2 > 0$ case

Let (x_i, y_i), $i = 1, 2, \ldots, n$ be n pairs of observations on (X, Y), x_{n+j}, $j = 1, \ldots, n_1$ and y_{n+n_1+k}, $k = 1, \ldots, n_2$ be n_1 and n_2 additional observations on X only and Y only respectively. Besides the notation (1), let

$$\overline{Y}^{(n_2)} = \frac{1}{n_2} \sum_{1}^{n_2} Y_{n+n_1+i}, \quad \overline{Y}^{(n+n_2)} = \frac{n\overline{Y}^{(n)} + n_2\overline{Y}^{(n_2)}}{n+n_2}.$$

Then Boas (1967) proposed an estimator of the covariance which uses the extra information from the single observations on X only and Y only. His estimator (C_2 in his notation) is given by

$$\hat{\gamma}_B = \frac{(n+n_1)(n+n_2)}{n[(n+n_1)(n+n_2)-n-n_1-n_2]} \sum_{i=1}^{n} (X_i - \overline{X}^{(n+n_1)})(Y_i - \overline{Y}^{(n+n_2)}). \quad \ldots (25)$$

We propose the following regression-estimator which is a generalization of $\hat{\gamma}_1$ considered in Section 3.

$$\hat{\gamma}_4 = \frac{\sum_{1}^{n} (X_i-\bar{X}^{(n)})(Y_i-\bar{Y}^{(n)})+\sum_{n+1}^{n+n_1}(X_i-\bar{X}^{(n_1)})(\hat{Y}_i-\bar{Y}^{(n_1)}) + \sum_{n+n_1+1}^{n+n_1+n_2}(Y_i-\bar{Y}^{(n_2)})(\hat{X}_i-\hat{\bar{X}}^{(n_2)}).}{n+n_1+n_2-3} \quad \ldots \quad (26)$$

The estimator $\hat{\gamma}_4$ is based on the estimated regression line (7) to estimate the missing values of Y. To estimate the missing values of X we use the regression line

$$\hat{X}_i = \bar{X}^{(n)}+ \frac{S_{12}}{S_{22}}(Y_i-\bar{Y}^{(n+n_2)}),\ i = n+n_1+1, \ldots, n+n_1+n_2. \quad \ldots \quad (27)$$

As in the case of $\hat{\gamma}_1$, the estimator of $\hat{\gamma}_4$ remains unchanged if we estimate μ_1 in (7) by $\bar{X}^{(n)}$ instead of $\bar{X}^{(n+n_1)}$ and μ_2 in (27) by $\bar{Y}^{(n)}$ instead of $\bar{Y}^{(n+n_2)}$. This estimator can be simplified to obtain

$$\hat{\gamma}_4 = \frac{1}{(n+n_1+n_2-3)}\Big[(n-1)S_{12}+\frac{S_{12}}{S_{11}}\sum_{n+1}^{n+n_1}(X_i-\bar{X}^{(n_1)})^2 + \frac{S_{12}}{S_{22}}\sum_{n+n_1+1}^{n+n_1+n^2}(Y_i-\bar{Y}^{(n_2)})^2\Big]. \quad \ldots \quad (28)$$

If we now assume that (X, Y) is distributed as bivariate normal we can write the variances of $\hat{\gamma}_B$ and $\hat{\gamma}_4$ as

$$V(\hat{\gamma}_B) = \sigma_1^2\sigma_2^2\Big[\frac{1+d}{n-1+dn}+\rho^2\frac{d^2+(n-1)(1+d)^2}{(n-1+dn)^2}\Big], \quad \ldots \quad (29)$$

and

$$V(\hat{\gamma}_4) = \frac{\sigma_1^2\sigma_2^2}{(n+n_1+n_2-3)^2}[(n+2n_1+2n_2-5)(1+\rho^2n)-4\rho^2(n_1+n_2-2) + \Delta(\rho)(n_1^2+n_2^2-2)+2(n_1-1)(n_2-1)E(R^2)]-\rho^2\sigma_1^2\sigma_2^2 \quad \ldots \quad (30)$$

where

$$d = \frac{n_1 n_2}{n(n+n_1+n_2)}, \text{ and } E(R^2) = 1-\frac{(n-2)}{(n-1)}(1-\rho^2)\,{}_2F_1\left(1, 1, \frac{n+1}{2}; \rho^2\right)$$

The function ${}_2F_1$ is the hypergeometric function.

To compare $\hat{\gamma}_B$ and $\hat{\gamma}_4$ numerically, relative efficiency of $\hat{\gamma}_B$ with respect to the regression-estimator $\hat{\gamma}_4$ viz. $e(\hat{\gamma}_B : \hat{\gamma}_4)$ was computed for $n = 8, 10, 15, 20$; $\rho^2 = 0$, $0{\cdot}1$ $(0{\cdot}2)0{\cdot}9$; $(\lambda_1, \lambda_2) = (0{\cdot}1, 0{\cdot}1)$, $(0{\cdot}1, 0{\cdot}2)$, $(0{\cdot}1, 0{\cdot}3)$, $(0{\cdot}1, 0{\cdot}4)$, $(0{\cdot}2, 0{\cdot}2)$ and $(0{\cdot}3, 0{\cdot}3)$, where $\lambda_1 = n_1/(n+n_1+n_2)$ and $\lambda_2 = n_2/(n+n_1+n_2)$. The resulting expression for $e(\hat{\gamma}_B : \hat{\gamma}_4)$ is symmetric in λ_1 and λ_2. Hence these computations also provided additional tables for values of $(\lambda_1, \lambda_2) = (0{\cdot}2, 0{\cdot}1)$, $(0{\cdot}3, 0{\cdot}1)$ and $(0{\cdot}4, 0{\cdot}1)$. There appears to be general loss in efficiency as ρ^2 increases for fixed n, λ_1, and λ_2 except in few cases where the gain is not significant. Also there appears to be general gain in efficiency as λ_1 (or λ_2) increases for fixed λ_2 (or λ_1), n and $\rho^2 \leqslant 0{\cdot}1$ while there is loss for $\rho^2 > 0{\cdot}1$. In Table 3 we present the value of $e(\hat{\gamma}_B : \hat{\gamma}_4)$ for $n = 10$ and 15 only.

TABLE 3. RELATIVE EFFICIENCY $e(\hat{\gamma}_B : \hat{\gamma}_4)$

$(n = 10)$

λ_1	λ_2	ρ^2					
		0.0	0.1	0.3	0.5	0.7	0.9
0.1	0.1	1.0160	1.0037	0.9849	0.9712	0.9608	0.9527
0.1	0.2	1.0576	1.0094	0.9360	0.8830	0.8435	0.8134
0.1	0.3	1.0952	1.0134	0.8888	0.7994	0.7332	0.6835
0.1	0.4	1.1298	1.0162	0.8435	0.7201	0.6294	0.5622
0.2	0.2	1.0884	1.0045	0.8782	0.7896	0.7263	0.6817
0.3	0.3	1.1418	0.9897	0.7641	0.6106	0.5068	0.4410
				$(n = 15)$			
0.1	0.1	1.0184	0.9942	0.9574	0.9316	0.9112	0.8962
0.1	0.2	1.0384	0.9852	0.9046	0.8474	0.8054	0.7743
0.1	0.3	1.0575	0.9761	0.8534	0.7666	0.7037	0.6576
0.1	0.4	1.0759	0.9672	0.8037	0.6886	0.6057	0.5457
0.2	0.2	1.0543	0.9701	0.8448	0.7583	0.6976	0.6558
0.3	0.3	1.0836	0.9379	0.7246	0.5825	0.4888	0.4310

It is observed that the regression-estimator $\hat{\gamma}_4$ is more efficient than $\hat{\gamma}_B$ for ρ^2 away from zero (except $\rho = 0$) and moderate n ($n = 15$ or greater). For small values of n ($n = 10$ or less), however $\hat{\gamma}_4$ is better than $\hat{\gamma}_B$ for $\rho^2 > 0{\cdot}1$ most of the time except in a few cases (when λ_1 and λ_2 are both small) and even there the loss in efficiency is not significant.

Acknowledgement. The authors wish to thank W. R. Javier for his numerical assistance and the referee for his critical review.

REFERENCES

ANDERSON, T. W. (1957): Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *J. Amer. Statist. Assoc.*, 52, 200-203.

BOAS, J. (1967): A note on the covariance between two random variables using extra information on the separate variables. *Statistica Neerlandica*, 21, 291-292.

GUPTA, A. K. and ROHATGI V. K. (1977): On a method of estimation from fragmentary samples. *Bull. Int. Statist. Inst.*, 47 (Book 4), 200-203.

——— (1978): Inference on the difference of means of correlated variables from fragmentary samples. *Sankhyā*, Series B, 40, 49-64.

OLKIN, I. and SYLVAN, M. (1977): Correlational analysis when some variances are covariances are known, *Mult. Analysis-IV*, ed. P. R. Krishnaiah, North-Holland, Amsterdam, 175-191.

Paper received : July, 1979.

Revised : September, 1981.