Survey Methodology, June 1987
Vol. 13, No. 1, pp. 109-121
Statistics Canada

Comparison of Estimators of Population Total in Two-Stage Successive Sampling Using Auxiliary Information

F.C. OKAFOR[1]

ABSTRACT

Singh and Srivastava (1973) proposed a linear unbiased estimator of the population mean when sampling on successive occasions using several auxiliary variables whose known population means remain unchanged for all occasions. In this paper, three composite estimators T_1, T_2 and T_3, each utilising an auxiliary variable whose known population mean changes from one occasion to the next, are presented for the estimation of the current population total. The proposed estimators are compared with the ordinary estimator, T_0, and the usual successive sampling estimator, T', of the current population total without the use of auxiliary information. We find that using auxiliary information in conjunction with successive sampling does not always uniformly produce a gain in efficiency over T_0 or T'. However, when applied to a survey of teak plantations to estimate the mean height of teak trees, T_1, T_2 and T_3 proved more efficient than T_0 and T'.

KEY WORDS: Successive occasion; Partial matching; Auxiliary variate.

1. INTRODUCTION

The theory and practice of surveying the same population at different points in time - technically called repetitive sampling or sampling over successive occasions - have been given considerable attention by some survey statisticians. The main objective of sampling on successive occasions is to estimate some population parameters (total, mean, ratio, etc) for the most recent occasion as well as changes in these parameters from one occasion to the next.

The theory of successive sampling was initiated by Jessen (1942). Many authors have since contributed, especially in the estimation of population means. Among them are Singh (1968), Abraham et al (1969), Kathuria and Singh (1971), and Kathuria (1975), to mention but a few.

Singh (1968) was the first to extend the theory of unistage sampling to two-stage sampling on successive occasions. He considered the sampling scheme in which, on the second occasion, a fraction λ of the first stage units (FSUs) selected on the previous occasion is retained, along with their selected second stage units (SSUs), and a fraction μ ($\lambda + \mu = 1$) selected afresh. He then obtained a minimum variance unbiased estimator of the population mean on the current occasion.

Abraham et al (1969) considered the situation in which partial matching of units was carried out at both stages. Units were selected by simple random sampling without replacement (SRSWOR). Kathuria (1975) modified this by using probability proportional to size and with replacement (PPSWR) for selection of the FSUs, and proposed a linear composite estimator for the population mean on the current occasion.

---

[1] F.C. Okafor, Department of Statistics, University of Ibadan, Ibadan, Nigeria.

When an auxiliary variable is highly correlated with the characteristic under study, the estimate of the population mean (total) of this characteristic can be improved using the auxiliary variable. Singh and Srivastava (1973) used auxiliary information to improve on the estimator of Singh (1968). They obtained a linear unbiased estimator of the population mean on the most recent occasion using several auxiliary variables whose population means are known and are the same for all occasions. Kathuria (1978) developed this study further by assuming that the population mean of the auxiliary variate is not known. He used a double sampling technique to estimate first the population mean of the auxiliary variate and then the mean of the characteristic under study.

In their contributions, Singh and Srivastava (1973) and Kathuria (1978) assumed that the necessary information on the auxiliary variables can be obtained from the respondents or reporting units (SSUs). This is not generally the case. It may happen that the information on the auxiliary variable is too distorted to be useful because of the sensitive nature of the question, or the respondents may refuse outright to supply any information. Alternatively, the information on the auxiliary variate may not be collected because the required question is not included in the questionnaire.

Singh and Srivastava also assumed that the known population total of the auxiliary variable is the same for all occasions. This may not be true in practice. If the population total of the main characteristic changes from one occasion to the next, there is every likelihood that the population total of any other variable correlated with it will also vary.

In this paper three composite estimators of the population total using auxiliary information and a two-stage successive sampling scheme are proposed. The performances of the three estimators are compared empirically and they are also applied to a survey of teak plantations to estimate the mean height of teak trees.

2. SAMPLING FOR TWO OCCASIONS

For all three proposed estimators, we assume that the population total of the auxiliary variable changes on the second occasion.

The estimators of the population total (mean) based on the partial matching scheme are better than the ordinary estimators of the population total (mean) without partial matching. Therefore, it is expected that the proposed estimators T_1, T_2 and T_3 will perform better than the ordinary population total estimator, T_o, and the estimator based on the partial matching scheme without the use of auxiliary information, T'.

In deriving these estimators, we assume that:

(i) the sample size is constant on each occasion;
(ii) the normed size measure P_i for the i^{th} first stage unit (FSU) is fixed for each occasion;
(iii) N and M_i, population sizes for the FSUs and the second stage units (SSUs) within the i^{th} FSU respectively, are constant for the two occasions;
(iv) the population total (mean) of the auxiliary variate is known.

Assumptions (i) – (iii) apply to T', T_1, T_2 and T_3; (iv) applies to T_1, T_2 and T_3, but not to T' and T_o.

On the first occasion, a sample S_1 of n FSUs is selected with probability proportional to size and with replacement (PPSWR) using P_i as normed size measure for the i^{th} $(i = 1, 2, \ldots, N)$ unit. For the selection of SSUs, we adopt the method due to Cochran

(1977, p. 306), which stipulates that if the i^{th} FSU in S_1 is drawn θ_i times ($i = 1, 2, \ldots, n$), we select θ_i independent subsamples of size m_i from the M_i SSUs.

On the second occasion, we select a sample of λn $(0 < \lambda < 1)$ FSUs from S_1 by simple random sampling without replacement (SRSWOR). The SSUs selected on the first occasion are retained for each of these λn matched FSUs. Then, a fresh sample of μn $(\mu = 1 - \lambda)$ FSUs is selected independently from the N FSUs by PPSWR, with P_i as normed size measure for the i^{th} FSU. In each of the μn FSUs, the SSUs are selected as on the first occasion.

3. NOTATION

We define y_{ij} (x_{ij}) as the value of the study variate for the j^{th} SSU in the i^{th} FSU on the current (previous) occasion. In addition, z_{hij} is defined as the value of the auxiliary variate for the j^{th} SSU in the i^{th} FSU on the h^{th} occasion ($h = 1, 2$). The sample means for SSUs in the i^{th} FSU are

$$\bar{x}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} x_{ij},\ \bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij} \text{ and } \bar{z}_{hi} = \frac{1}{m_i} \sum_{j=1}^{m_i} z_{hij}.$$

The population total for the i^{th} FSU and the overall population total for the auxiliary variate are

$$Z_{hi} = \sum_{j=1}^{M_i} z_{hij} \text{ and } Z_h = \sum_{j=1}^{N} Z_{hi}.$$

We define additional notation as follows:

$$S_b^2\,(y) = \sum_{i=1}^{N} P_i \left(\frac{Y_i}{P_i} - Y\right)^2 \text{ is the between - FSU variance;}$$

$$S_w^2\,(y) = \sum_{i=1}^{N} \frac{M_i^2}{P_i} \left(\frac{1}{m_i} - \frac{1}{M_i}\right) S_{wi}^2\,(y) \text{ is the variance among SSUs within the FSUs;}$$

$$S_{wi}^2\,(y) = \frac{1}{M_i - 1} \sum_{j=1}^{M_i} (y_{ij} - \bar{y}_i)^2 \text{ is the variance among the SSUs in the } i^{th} \text{ FSU;}$$

$$S^2\,(y) = S_b^2\,(y) + S_w^2\,(y);$$

$$C_b\,(x,y) = \rho_b S_b(x)\, S_b(y) \text{ is the between-FSU covariance of } x \text{ and } y;$$

$$C_w(x,y) = \rho_w S_w(x)\, S_w(y) \text{ is the covariance of } x \text{ and } y \text{ among SSUs within the FSUs;}$$

$$C(x,y) = C_b(x,y) + C_w(x,y).$$

The between- and within-FSU correlation coefficients between x and y are respectively ρ_b and ρ_w.

4. ESTIMATORS FOR THE POPULATION TOTAL AND THEIR OPTIMUM VARIANCES

4.1 Case (i)

The first estimator of the population total, Y, on the second occasion is used when information on the auxiliary variable is not available but the FSU population total of the auxiliary variable is available for the selected FSUs. It is given as

$$T_1 = \theta(1)\, T_m(1) + (1 - \theta(1))\, T_u(1) \tag{4.1}$$

$\theta(1)$ is a constant chosen so that the variance of T_1, $V(T_1)$, attains a minimum; while

$$\begin{aligned} T_m(1) = {} & \frac{1}{\lambda n} \sum_{i=1}^{\lambda n} \left\{ \frac{M_i \bar{y}_i}{P_i} - k(1) \left(\frac{Z_{2i}}{P_i} - Z_2 \right) \right\} \\ & - b(1) \left[\frac{1}{\lambda n} \sum_{i=1}^{\lambda n} \left\{ \frac{M_i \bar{x}_i}{P_i} - k(1) \left(\frac{Z_{1i}}{P_i} - Z_1 \right) \right\} \right. \\ & \left. - \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{M_i \bar{x}_i}{P_i} - k(1) \left(\frac{Z_{1i}}{P_i} - Z_1 \right) \right\} \right] \end{aligned}$$

is the difference estimator of Y based on the matched sample;

$$T_u(1) = \frac{1}{n\mu} \sum_{i=1}^{n\mu} \left\{ \frac{M_i \bar{y}_i}{P_i} - k(1) \left(\frac{Z_{2i}}{P_i} - Z_2 \right) \right\}$$

is the estimator for Y based on the unmatched sample; and $k(1)$ and $b(1)$ are known constants.

For this estimator, it is assumed that the population total of the auxiliary variate, Z_i, is available for each selected FSU on each occasion. The overall population total, Z, is also available on each occasion. No additional information on the auxiliary variate is obtained from the respondents (SSUs).

Now by minimizing $V(T_1)$ with respect to $\theta(1)$ and solving, the optimum value of $\theta(1)$ becomes

$$\theta_0(1) = \lambda A_2(1) / \Delta(1)$$

where

$$A_2(1) = S^2(y) + k^2(1)\, S_b^2(z_2) - 2k(1)\, C_b(z_2, y),$$

$$\Delta(1) = A_2(1) + \mu^2 \{ b^2(1)\, A_1(1) - 2b(1) \beta(1) \}.$$

The optimum value of $k(1)$ is obtained by minimizing $V(T_u(1))$ with respect to $k(1)$. This gives $k_0(1) = C_b(z_2, y) / S_b^2(z_2)$.

It can be shown that the optimum $V(T_1)$ for a given λ, following the method adopted by Jessen (1942), is

$$V_0(T_1) = \frac{1}{n}[A_2(1) + \mu\{b^2(1)A_1(1) - 2b(1)\beta(1)\}]A_2(1)/\Delta(1) \qquad (4.2)$$

where

$$A_1(1) = S^2(x) + k^2(1)S_b^2(z_1) - 2k(1)C_b(z_1,x),$$

$$\beta(1) = C(x,y) + k^2(1)C_b(z_1,z_2) - k(1)\{C_b(x,z_2) + C_b(z_1,y)\},$$

$$\Delta(1) = A_2(1) + \mu^2\{b^2(1)A_1(1) - 2b(1)\beta(1)\}.$$

Minimizing the variance of $T_m(1)$, the optimum $b(1)$ is

$$b_0(1) = \beta(1)/A_1(1).$$

If $b_0(1)$ is substituted in (4.2), the optimum variance becomes

$$V_0(T_1) = \frac{1}{n}\left[\frac{A_1(1)A_2(1) - \mu\beta^2(1)}{A_1(1)A_2(1) - \mu^2\beta^2(1)}\right]A_2(1). \qquad (4.3)$$

By minimizing $V_o(T_1)$ in (4.2) with respect to μ, the optimum matching fraction boils down to $\lambda_0 = 1 - \mu_0$ where

$$\mu_0 = A_2(1)[A_2(1) + \{A_2^2(1) + A_2(1)(b^2(1)A_1(1) - 2b(1)\beta(1))\}^{1/2}]^{-1}. \qquad (4.4)$$

If $A_2(1) = A_1(1)$, i.e. the population variability is the same on both occasions, the expression in (4.3) yields

$$V_0(T_1) = \frac{1}{n}\left[\frac{A^2(1) - \mu\beta^2(1)}{A^2(1) - \mu^2\beta^2(1)}\right]A(1) \qquad (4.5)$$

while the optimum matching fraction, μ_0 (given in (4.4)), with $b_0(1)$ substitued for $b(1)$ becomes

$$\mu_0 = A(1)[A(1) + \{A^2(1) - \beta^2(1)\}^{1/2}]^{-1}. \qquad (4.6)$$

When μ_0 is substituted in (4.5) the variance works out as

$$V_0(T_1) = \frac{1}{2n}[A(1) + \{A^2(1) - \beta^2(1)\}^{1/2}]. \qquad (4.7)$$

4.2 Case (ii)

The second estimator is the usual one in which information is obtained on both the main and auxiliary characteristic from the reporting units and the population total of the auxiliary characteristic is known.

It is written as

$$T_2 = \theta(2)\, T_m(2) + (1 - \theta(2))\, T_u(2), \tag{4.8}$$

where

$$\begin{aligned} T_m(2) = \frac{1}{\lambda n} \sum_{i=1}^{\lambda n} & \left\{ \frac{M_i \bar{y}_i}{P_i} - k(2) \left(\frac{M_i \bar{z}_{2i}}{P_i} - Z_2 \right) \right. \\ & - b(2) \left[\frac{1}{\lambda n} \sum_{i=1}^{\lambda n} \left\{ \frac{M_i \bar{x}_i}{P_i} - k(2) \left(\frac{M_i \bar{z}_{1i}}{P_i} - Z_1 \right) \right\} \right. \\ & \left. - \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{M_i \bar{x}_i}{P_i} - k(2) \left(\frac{M_i \bar{z}_{1i}}{P_i} - Z_1 \right) \right\} \right], \end{aligned}$$

and

$$T_u(2) = \frac{1}{n\mu} \sum_{i=1}^{n\mu} \left\{ \frac{M_i \bar{y}_i}{P_i} - k(2) \left(\frac{M_i \bar{z}_{2i}}{P_i} - Z_2 \right) \right\}.$$

Here the overall population total of the auxiliary variate is known on both occasions. In addition, information on the auxiliary variate, z_{ij}, is obtained for every SSU in the sample. This is the usual way of using the auxiliary information in a two- stage design described in the literature. It can be shown that the optimum variance of T_2 is

$$V_0(T_2) = \frac{1}{n} [A_2(2) + \mu\{b^2(2) A_1(2) - 2b(2)\beta(2)\}]\, A_2(2)/\Delta(2) \tag{4.9}$$

and the optimum weight is

$$\theta_0(2) = \lambda A_2(2)/\Delta(2)$$

where

$$\begin{aligned} A_2(2) &= S^2(y) + k^2(2)\, S^2(z_2) - 2k(2)\, C(z_2, y), \\ A_1(2) &= S^2(x) + k^2(2)\, S^2(z_1) - 2k(2)\, C(z_1, x), \\ \beta(2) &= C(x, y) + k^2(2)\, C(z_1, z_2) - k(2)\, \{C(z_1, y) + C(x, z_2)\}, \\ \Delta(2) &= A_2(2) + \mu^2\, \{b^2(2)\, A_1(2) - 2b(2)\, \beta(2)\}. \end{aligned}$$

The optimum value of $k(2)$ is $k_0(2) = C(z_2,y)/S_2(z_2)$.

By substituting the optimum regression coefficient $b_0(2) = \beta(2)/A_1(2)$, obtained by minimizing the variance of $T_m(2)$, in (4.9) and assuming that $A_2(2) = A_1(2) = A(2)$ we have

$$V_0(T_2) = \frac{1}{n}\left[\frac{A^2(2) - \mu\beta^2(2)}{A^2(2) - \mu^2\beta^2(2)}\right] A(2). \tag{4.10}$$

If the optimum μ is substituted in (4.10), the variance becomes

$$V_0(T_2) = \frac{1}{2n}\,[A(2) + \{A^2(2) - \beta^2(2)\}^{1/2}]. \tag{4.11}$$

4.3 Case (iii)

The third way of utilising available auxiliary information to improve the estimate of the current population total, Y, under the given sampling scheme is similar to the second. The only difference is that the population total of the auxiliary characteristic is not known; however, its FSU population mean is known for the selected FSUs.

This is given as

$$T_3 = \theta(3)\, T_m(3) + (1 - \theta(3))\, T_u(3), \tag{4.12}$$

where

$$\begin{aligned} T_m(3) = {} & \frac{1}{\lambda n}\sum_{i=1}^{\lambda n}\frac{M_i}{P_i}\{\bar{y}_i - k(3)\,(\bar{z}_{2i} - \bar{Z}_{2i})\} \\ & - b(3)\left[\frac{1}{\lambda n}\sum_{i=1}^{\lambda n}\frac{M_i}{P_i}\{\bar{x}_i - k(3)\,(\bar{z}_{1i} - \bar{Z}_{1i})\}\right. \\ & \left. - \frac{1}{n}\sum_{i=1}^{n}\frac{M_i}{P_i}\{\bar{x}_i - k(3)\,(\bar{z}_{1i} - \bar{Z}_{1i})\}\right], \end{aligned}$$

and

$$T_u(3) = \frac{1}{n\mu}\sum_{i=1}^{n\mu}\frac{M_i}{P_i}\{\bar{y}_i - k(3)\,(\bar{z}_{2i} - \bar{Z}_{2i})\}.$$

For this estimator, we suppose that the values of both the main variate and the auxiliary variate are obtained for every SSU in the sample on both occasions. We also assume that the population mean, $\bar{Z}_i$, of the auxiliary variate is known for the selected FSUs.

The optimum variance of T_3 for a given λ is given as

$$V_0(T_3) = \frac{1}{n}\,[A_2(3) + \mu\{b^2(3)A_1(3) - 2b(3)\beta(3)\}]\,A_2(3)/\Delta(3) \tag{4.13}$$

while the optimum weight is as usual obtained as

$$\theta_0(3) = \lambda A_2(3)/\Delta(3),$$

where

$$A_2(3) = S^2(y) + k^2(3)\; S^2_w(z_2) - 2k(3)\; C_w(z_2,y),$$

$$A_1(3) = S^2(x) + k^2(3)\; S^2_w(z_1) - 2k(3)\; C_w(z_1,x),$$

$$\beta(3) = C(x,y) + k^2(3)\; C_w(z_1,z_2) - k(3)\; \{C_w(z_1,y) + C_w(z_2,x)\},$$

$$\Delta(3) = A_2(3) + \mu^2\; \{b^2(3)A_1(3) - 2b(3)\; \beta(3)\}.$$

The optimum value of $k(3)$ is $k_0(3) = C_w(z_2,y)/S^2_w(z_2)$.

If the optimum regression coefficient is substituted in (4.13), and it is assumed that population variances are the same on both occasions, then (4.13) works out as

$$V_0(T_3) = \frac{1}{n}\left[\frac{A^2(3) - \mu\beta^2(3)}{A^2(3) - \mu^2\beta^2(3)}\right] A(3). \tag{4.14}$$

When the optimum μ is substituted in (4.14), the variance is

$$V_0(T_3) = \frac{1}{2n}\; [A(3) + \{A^2(3) - \beta^2(3)\}^{1/2}]. \tag{4.15}$$

4.4 Efficiency of the Proposed Estimators

The variances given in (4.7), (4.11) and (4.15) will be used to compare the efficiencies of T_1, T_2 and T_3 with respect to

$$T_0 = \frac{1}{n}\sum_{i=1}^{n}\frac{M_i\bar{y}_i}{P_i}.$$

T_0 is the estimator for y when there is no partial matching of units and no auxiliary information used. In addition, the efficiency of T_0 compared to the usual partial matching estimator T', which uses no auxiliary information, will be presented to assist in understanding the performance of the proposed estimators.

The usual partial matching estimator is defined as

$$T' = \theta' T'_m + (1 - \theta')\, T'_u, \tag{4.16}$$

where

$$T'_m = \frac{1}{\lambda n} \sum_{i=1}^{\lambda n} \frac{M_i \bar{y}_i}{P_i} - b' \left\{ \frac{1}{\lambda n} \sum_{i=1}^{\lambda n} \frac{M_i \bar{x}_i}{P_i} - \frac{1}{n} \sum_{i=1}^{n} \frac{M_i \bar{x}_i}{P_i} \right\},$$

and

$$T'_u = \frac{1}{n\mu} \sum_{i=1}^{n\mu} \frac{M_i \bar{y}_i}{P_i}.$$

The optimum variance of T', obtained using the optimum value of b', $b'_0 = C(x,y)/S^2(x)$, and assuming $S^2(y) = S^2(x)$ is

$$V_0(T') = \frac{1}{n} \left[\frac{S^2(y) - \mu C(x,y)}{S^2(y) - \mu^2 C(x,y)} \right] S^2(y). \tag{4.17}$$

Substituting the optimum value of μ in (4.17), the variance of T' becomes

$$V_0(T') = \frac{1}{2n} [S^2(y) + \{S^4(y) - C^2(x,y)\}^{1/2}]. \tag{4.18}$$

To calculate the efficiencies, the following assumptions about the correlation coefficients and the constant k were made:

$$\rho_b(x,z_2) = \rho_b(z_1,y) = \rho_b(z_1,z_2) = \rho_b;$$

$$\rho_w(x,z_2) = \rho_w(z_1,y) = \rho_w(z_1,z_2) = \rho_w;$$

$$k(1) = k(2) = k(3) = 1.$$

The efficiencies have been presented for only the positive values of ρ_b and ρ_w, and a set of values of

$$\delta = S_w^2(y)/S_b^2(y), \; R_b = S_b^2(z)/S_b^2(y) \text{ and } R_w = S_w^2(z)/S_b^2(y).$$

Looking at Table 2, we observe that none of the strategies T_1, T_2 and T_3 (sampling design and estimator) is uniformly more efficient than strategy T_0. The contrary is true of T', which is always more efficient than T_0; at worst, its gain over T_0 is small (see Table 1).

The results in Tables 1 and 2 show T_1 is to be preferred to T' only when $R_b = 0.05$; and when $\rho_b = 0.8$ and $R_b = 0.5$.

Table 1

The Efficiency of T' with Respect to T_0

ρ_b	δ	$\rho_w = 0.2$	$\rho_w = 0.8$
0.2	0.05	1.01	1.01
	0.5	1.01	1.04
	5.0	1.01	1.17
	0.05	1.22	1.25
0.8	0.5	1.11	1.25
	5.0	1.02	1.25

T_2 is better than T' when:

(i) $\rho_w = 0.2,\ R_b = R_w = 0.05;$

(ii) $\rho_b = \rho_w = 0.8,\ R_b = R_w = 0.05,\ 0.5;$

(iii) $\delta = 0.5,\ 5.0,\ R_w = R_b = 0.05,\ \rho 0.5,\ \rho_b = 0.2$ and $\rho_w = 0.8$.

T_3 is generally more efficient than T' when:

(i) $\delta = 5.0,\ \rho_w = 0.8;$

(ii) $\delta = 0.5,\ \rho_w = 0.8$ and $R_w = 0.05,\ 0.5$.

The maximum gain in efficiency of T' over T_0 is 25% (see Table 1). In Table 2, the maximum gain of T_1 over T_0 is 155%, which occurs when $\rho_b = \rho_w = 0.8, \delta = 0.05, R_b = 0.5$. The maximum gain in efficiency of T_2 over T_0 is 172%; this happens when $\rho_b = \rho_w = 0.8, \delta = R_w = 0.05$. We also observe that when $\rho_b = \rho_w = 0.8, \delta = R_w = 5.0$, the maximum gain of T_3 over T_0 is 104%. It is therefore evident that the use of an auxiliary variate has tremendously improved the efficiency of partial matching of units.

If we now take the three strategies T_1, T_2 and T_3, and compare them among themselves, we conclude that none of the strategies is uniformly better than the other, even though the maximum gain in efficiency of T_2 over T_0 is higher than that of T_1, which in turn is higher than the maximum gain of T_3. In general T_1 is superior to T_2 when $\rho_w = 0.2$, while T_2 is better than T_1 when $\rho_w = 0.8$. T_1 is preferred to T_3 when $\rho_b = 0.8$, $\rho_w = 0.2$ and $R_b = 0.05, 0.5$, and also when $\rho_b = \rho_w = 0.8$ and $\delta = R_b = 0.05$. Finally T_3 is better than T_2 when $\rho_w = 0.8$, $R_b = 5.0$, and when $\rho_b = \rho_w = 0.2$ with $R_b = 0.5, 5.0$.

5. APPLICATION

The proposed estimators were applied to a survey of teak plantations. The aim was to estimate the average height of teak trees using the girth as the auxiliary information.

Table 2

The Efficiency of T_1, T_2, and T_3 with Respect to T_0

		$\rho_w = 0.2$								
		$R_b = 0.05$			$R_b = 0.5$			$R_b = 5.0$		
		R_w			R_w			R_w		
ρ_b	δ	0.05	0.5	5.0	0.05	0.5	5.0	0.05	0.5	Strategy
0.2	0.05	1.04	1.04	1.04	0.83	0.83	0.83	0.20	0.20	T_1
		1.01	0.73	0.18	0.81	0.62	0.17	0.20	0.19	T_2
		0.98	0.71	0.18	0.98	0.71	0.18	0.98	0.71	T_3
	0.5	1.03	1.03	1.03	0.87	0.87	0.87	0.27	0.27	T_1
		1.04	0.85	0.26	0.88	0.74	0.25	0.27	0.25	T_2
		1.02	0.84	0.26	1.02	0.84	0.26	1.02	0.84	T_3
	5.0	1.02	1.02	1.02	0.97	0.97	0.97	0.60	0.60	T_1
		1.04	1.03	0.67	0.99	0.99	0.65	0.60	0.60	T_2
		1.03	1.03	0.67	1.03	1.03	0.67	1.03	1.03	T_3
0.8	0.05	1.62	1.62	1.62	2.53	2.53	2.53	0.45	0.45	T_1
		1.53	0.94	0.19	2.35	1.23	0.20	0.45	0.38	T_2
		1.16	0.77	0.18	1.16	0.77	0.18	1.16	0.77	T_3
	0.5	1.34	1.34	1.34	1.74	1.74	1.74	0.45	0.45	T_1
		1.34	1.03	0.27	1.76	1.28	0.29	0.54	0.48	T_2
		1.11	0.88	0.26	1.11	0.88	0.26	1.11	0.88	T_3
	5.0	1.07	1.07	1.07	1.13	1.13	1.13	0.83	0.83	T_1
		1.10	1.09	0.69	1.16	1.15	0.72	0.84	0.83	T_2
		1.05	1.03	0.67	1.05	1.03	0.67	1.05	1.03	T_3

			$\rho_w = 0.8$									
			$R_b = 0.05$			$R_b = 0.5$			$R_b = 5.0$			
			R_w			R_w			R_w			
ρ_b	δ	5.0	0.05	0.5	5.0	0.05	0.5	5.0	0.05	0.5	5.0	Strategy
0.2	0.05	0.20	1.05	1.05	1.05	0.83	0.83	0.83	0.20	0.20	0.20	T_1
		0.11	1.07	0.85	0.23	0.85	0.70	0.21	0.19	0.19	0.12	T_2
		0.18	1.04	0.83	0.23	1.04	0.83	0.23	1.04	0.83	0.23	T_3
	0.5	0.27	1.06	1.06	0.89	0.89	0.89	0.89	0.27	0.27	0.27	T_1
		0.15	1.21	1.30	0.41	1.00	1.06	0.38	0.28	0.28	0.19	T_2
		0.26	1.18	1.26	0.41	1.18	1.26	0.41	1.18	1.26	0.41	T_3
	5.0	0.60	1.17	1.17	1.17	1.09	1.09	1.09	0.62	0.62	0.62	T_1
		0.46	1.31	1.64	2.03	1.22	1.51	1.87	0.67	0.76	0.84	T_2
		0.67	1.30	1.63	2.00	1.30	1.63	2.00	1.30	1.63	2.00	T_3
0.8	0.05	0.45	1.65	1.65	1.65	2.55	2.55	2.55	0.46	0.46	0.46	T_1
		0.15	1.70	1.22	0.25	2.72	1.64	0.27	0.46	0.42	0.18	T_2
		0.18	1.27	0.98	0.24	1.26	0.98	0.24	1.27	0.98	0.24	T_3
	0.5	0.45	1.50	1.50	1.50	1.88	1.88	1.88	0.56	0.56	0.56	T_1
		0.21	1.75	1.83	0.46	2.34	2.65	0.50	0.59	0.61	0.31	T_2
		0.26	1.40	1.43	0.43	1.40	1.43	0.43	1.40	1.43	0.43	T_3
	5.0	0.83	1.30	1.30	1.30	1.35	1.35	1.35	0.95	0.95	0.95	T_1
		0.85	1.46	1.85	2.25	1.53	1.98	2.53	1.03	1.22	1.38	T_2
		0.67	1.39	1.74	2.04	1.39	1.74	2.04	1.39	1.74	2.04	T_3

Table 3

Estimated Efficiency of the Proposed Estimators with Respect to T_0 in the Estimation of the Average Height of Teak Trees

Estimators	Mean height (m)	Variance (m^2)	Estimated % Efficiency
T_0 (no matching)	20.04	6.3118	100
T' Partial matching	18.06	4.0680	155
T_1	17.86	0.0718	8791
T_2	17.31	0.0651	9635
T_3	17.99	4.0183	157

The teak trees used in this study were planted in 1965 with different spacings, producing plantations with the following number of trees per hectare: 2,000, 800, 400 and 250 trees. To measure the trees, an area of 40 metres by 40 metres was mapped out in each plantation after a sample of 8 plantations (FSUs) had been selected from 16 plantations, using the PPSWR scheme. The number of trees in each plantation was used as a measure of size. All the trees in the 40m by 40m area constituted the second stage units and the girth of each tree at breast height was measured. For the height measurements, a subsample of the trees was selected from the 40m by 40m area in each selected FSU. The first measurements were carried out in 1981 and the second in 1983. The sampling scheme used was the same as the one described in Section 2, with 50% matching of the FSUs.

The estimated efficiencies are given in Table 3. The sample estimates of the variance and covariance terms were used to obtain the optimum variances of T', T_1, T_2 and T_3 because the population values of these variances and covariances were not known. Therefore, the low values of the estimated optimum variances of T_1 and T_2 can be attributed partly to the nature of the sample data and partly to the nature of the estimators.

We observe that the estimator T_2 is more efficient than either T_1 or T_3, while T_1 is more efficient than T_3 in the estimation of the average height of teak trees using the girth as the auxiliary information.

ACKNOWLEDGEMENTS

I am grateful to the referee and the associate editor for their useful comments for the improvement of this paper. I thank Dr. O. Abe of the Department of Statistics, University of Ibadan, Ibadan, for going over the draft of the revised paper.

REFERENCES

ABRAHAM, T.P., KHOSLA, R.K., and KATHURIA, O.P. (1969). Some investigations of the use of successive sampling in pest and disease surveys. *Journal of the Indian Society of Agricultural Statistics*, 21, 43 - 57.

COCHRAN, W.G. (1977). *Sampling Techniques*, (3rd. ed.). New York: John Wiley.

JESSEN, R.J. (1942). Statistical investigations of a sample survey for obtaining farm facts. *Iowa Agricultural Experimental Station Research Bulletin*, 304, 54-59.

KATHURIA, O.P. (1975). Some estimators in two-stage sampling on successive occasions with partial matching at both stages. *Sankhyā*, Ser. C, 37, 147 - 162.

KATHURIA, O.P. (1978). Double sampling on successive occasions using a two-stage design. *Journal of the Indian Society of Agricultural Statistics*, 30, 49 - 64.

KATHURIA, O.P., and SINGH, D. (1971). Relative efficiencies of some alternative procedures in two-stage sampling on successive occasions. *Journal of the Indian Society of Agricultural Statistics*, 23, 101 - 114.

SINGH, S., and SRIVASTAVA, A.K. (1973). Use of auxiliary information in two-stage successive sampling. *Journal of the Indian Society of Agricultural Statistics*, 25, 101 - 114.

SINGH, D. (1968). Estimates in successive sampling using a multistage design. *Journal of the American Statistical Association*, 63, 99 - 112.