Hacettepe Journal of Mathematics and Statistics
Volume 39 (1) (2010), 121 – 129

ESTIMATION OF THE FINITE POPULATION MEAN IN TWO PHASE SAMPLING WHEN AUXILIARY VARIABLES ARE ATTRIBUTES

Javid Shabbir*[†] and Sat Gupta[‡]

Received 16 : 06 : 2009 : Accepted 18 : 09 : 2009

Abstract

We consider the problem of estimating the finite population mean in two phase sampling when some information on auxiliary attributes is available. It is shown that the proposed estimator is more efficient than the usual mean estimator, the S. Bahl and R. K. Tuteja (*Ratio and product type exponential estimators*, Information and Optimization Sciences **12** (1), 159–163, 1991) estimator and a family of estimators considered by H. S. Jhajj *et al* (*A family of estimators of population mean using information on auxiliary attribute*, Pakistan Journal of Statistics **22** (1), 43–50, 2006). Two numerical examples are considered to further evaluate the performances of various estimators considered here.

Keywords: Two-phase sampling, Attribute, Point bi-serial correlation, Phi correlation, Efficiency.

2000 AMS Classification: 62 D O5.

1. Introduction

There are many situations when auxiliary information is available in the form of attributes. For example, sex is a good auxiliary attribute while dealing with height, and the breed of a cow is a good auxiliary attribute while estimating milk production (see Naik and Gupta [9]). Another example considered by Jhajj *et al.* [5] is where crop variety is used as an auxiliary attribute in estimating the yield of wheat. In all of these examples point bi-serial correlation between the study variable and the auxiliary attribute exists. Naik and Gupta [9] introduce a ratio estimator when the study variable and the auxiliary attribute are positively correlated. Jhajj *et al.* [5] discuss a family of

*Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan. E-mail: jsqau@yahoo.com

[†]Corresponding Author.

[‡]Department of Mathematics and Statistics, University of North Carolina at Greensboro, Greensboro, NC 27412, USA. E-mail: sngupta@uncg.edu

estimators for the population mean in single and two-phase sampling when the study variable and the auxiliary attribute are positively correlated. We extend the works of Naik and Gupta [9] and Jhajj et al. [5] to the situation when two auxiliary attributes are available. We assume that both auxiliary attributes have significant point bi-serial correlation with the study variable and that there is significant phi-correlation between the two auxiliary attributes. Two examples of this scenario are:

(a) The study variable (y) is the yield of some crop in a particular region, ψ_1 is the ownership (rich vs. poor) of the cultivated land and ψ_2 is the irrigation status (adequate vs. inadequate).

(b) The study variable (y) is the student's grade point average, ψ_1 is the number of hours spent studying (low vs. high) and ψ_2 is the use of library facilities (low vs. high).

In the examples above we expect point bi-serial correlation between the study variable and the two auxiliary attributes; and phi-correlation between the two auxiliary attributes.

Now consider a finite population which consists of N identifiable units U_i $(1 \leq i \leq N)$. Assume there is a complete dichotomy in the population regarding the presence and absence of the two attributes ψ_j, $(j = 1, 2)$, which take values zero or one. Let y_i and ψ_{ji} $(j = 1, 2)$ be the observations on the study variable y and the auxiliary attributes ψ_j, $(j = 1, 2)$, respectively. Let $\psi_{ji} = 1$, if i^{th} unit possesses the attribute ψ_j, $(j = 1, 2)$ and $\psi_{ji} = 0$ otherwise. Let $A_j = \sum_{i=1}^{N} \psi_{ji}$ and $a_j = \sum_{i=1}^{n} \psi_{ji}$ denote the total number of units in the population and the sample, respectively, that possess the attribute ψ_j. Let the corresponding population and sample proportions be $P_j = \frac{\sum_{i=1}^{N} \psi_{ji}}{N} = \frac{A_j}{N}$ and $p_j = \frac{\sum_{i=1}^{n} \psi_{ji}}{n} = \frac{a_j}{n}$, $(j = 1, 2)$ respectively. Let $s_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$ and $s_{\psi_j}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\psi_{ji} - p_j)^2$ be the sample variances corresponding to the population variances $S_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2$ and $S_{\psi_j}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\psi_{ji} - P_j)^2$, respectively, where $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ and $\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i$.

Let $s_{y\psi_j} = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})(\psi_{ji} - p_j)$ and $\hat{\rho}_{pb_j} = \frac{s_{y\psi_j}}{s_y s_{\psi_j}}$, $(j = 1, 2)$ be the sample point bi-serial covariance and point bi-serial correlation between y and ψ_j corresponding to the population point bi-serial covariance and point bi-serial correlation $S_{y\psi_j} = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})(\psi_{ji} - P_j)$ and $\rho_{pb_j} = \frac{S_{y\psi_j}}{S_y S_{\psi_j}}$, respectively.

Let $s_{\psi_1\psi_2} = \frac{1}{n-1} \sum_{i=1}^{n} (\psi_{1i} - p_1)(\psi_{2i} - p_2)$ and $\hat{\rho}_{\phi} = \frac{s_{\psi_1\psi_2}}{s_{\psi_1} s_{\psi_2}}$ be the sample phi-covariance and phi correlation between ψ_1 and ψ_2 corresponding to the population phi-covariance and phi correlation $S_{\psi_1\psi_2} = \frac{1}{N-1} \sum_{i=1}^{N} (\psi_{1i} - P_1)(\psi_{2i} - P_2)$ and $\rho_{\phi} = \frac{S_{\psi_1\psi_2}}{S_{\psi_1} S_{\psi_2}}$.

Let $C_y = \frac{S_y}{\bar{Y}}$ and $C_{p_j} = \frac{S_{\psi j}}{P_j}$ $(j = 1, 2)$ be the coefficients of variation of y and ψ_j, $(j = 1, 2)$, respectively.

It is assumed that population proportion (P_1) and population variance ($S_{\psi_1}^2$) for the first auxiliary attribute ψ_1 are unknown, while the same is known for the second auxiliary attribute ψ_2. In such a situation, we can estimate (P_1) and ($S_{\psi_1}^2$) from the sample by using a two-phase sampling procedure as per Jhajj et al. [6], Kiregyra [7] and Swain [15]. We use simple random sampling without replacement at both phases as described below:

(1) We draw a sample s^* of fixed size n^* from the population and observe $(\psi_j, j = 1, 2)$ and estimate P_1 as well as $S_{\psi_1}^2$.

(2) Given s^* , we draw a sample s $(s \subset s^*)$ of fixed size n and observed y.

Let $p_j^* = \frac{1}{n^*} \sum_{i=1}^{n^*} \psi_{ji}$, $(j = 1, 2)$, $s_{\psi_j}^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (\psi_{ji} - p_j)^2$, $s_{\psi_j}^{*2} = \frac{1}{(n^*-1)} \sum_{i=1}^{n^*} (\psi_{ji} - p_j^*)^2$.

We now introduce the following relative error terms. Let

$$\Delta_0 = \frac{\bar{y}-\bar{Y}}{\bar{Y}}, \qquad \Delta_{\psi_1} = \frac{p_1 - P_1}{P_1}, \qquad \Delta^*_{\psi_1} = \frac{p^*_1 - P_1}{P_1},$$
$$\Delta_{\psi_2} = \frac{s^2_{\psi_1} - S^2_{\psi_1}}{S^2_{\psi_1}}, \qquad \Delta^*_{\psi_2} = \frac{s^{*2}_{\psi_1} - S^2_{\psi_1}}{S^2_{\psi_1}}, \qquad \Delta_{\psi_3} = \frac{p_2 - P_2}{P_2},$$
$$\Delta^*_{\psi_3} = \frac{p^*_2 - P_2}{P_2}, \qquad \Delta_{\psi_4} = \frac{s^2_{\psi_2} - S^2_{\psi_2}}{S^2_{\psi_2}}, \qquad \Delta^*_{\psi_4} = \frac{s^{*2}_{\psi_2} - S^2_{\psi_2}}{S^2_{\psi_2}}.$$

Expected values of these relative errors are calculated in the Appendix.

We also introduce two more notations:

$$\mu_{abc} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^a(\psi_{1i} - P_1)^b(\psi_{2i} - P_2)^c \text{ and } \delta_{abc} = \frac{\mu_{abc}}{\mu_{200}^{a/2}\,\mu_{020}^{b/2}\,\mu_{002}^{c/2}}.$$

We now discuss some of the existing estimators of μ.

The usual estimator of μ is $\bar{y}$, and its variance is given by

$$\text{(1.1)} \qquad \operatorname{Var}(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right) S^2_y.$$

The class of estimators using a single auxiliary attribute in two phase sampling when P_1 is unknown is discussed by Jhajj *et al.* [5]. It is given by

$$\text{(1.2)} \qquad \bar{y}_J = h(\bar{y}, \upsilon^*),$$

where $\upsilon^* = \frac{p_1}{p^*_1}$ and $h(\bar{y}, \upsilon^*)$ is a parametric function of $\bar{y}$ and υ^* such that $h(\bar{Y}, 1) = \bar{Y}$, $\forall\,\bar{Y}$, and certain regularity conditions hold (see Jhajj *et al.* [5]). The minimum MSE of $\bar{y}_J$ to the first order of approximation, is given by

$$\text{(1.3)} \qquad \operatorname{MSE}(\bar{y}_J)_{\min} \cong \left(\frac{1}{n} - \frac{1}{N}\right) S^2_y - \left(\frac{1}{n} - \frac{1}{n^*}\right) S^2_y \rho^2_{pb_1}.$$

The above expression is equal to the variance of the linear regression estimator $\bar{y}_{lrd} = \bar{y} + b(p^*_1 - p_1)$ in two phase sampling, where $b = \frac{\hat{\rho}_{pb_1} s_y}{s_{\psi_1}}$ is the sample regression coefficient of y on ψ_1.

The Bahl and Tuteja [2] estimator using a single auxiliary attribute in two phase sampling when P_1 is unknown, is given by

$$\text{(1.4)} \qquad \bar{y}_{BT} = \bar{y}\exp\left(\frac{p^*_1 - p_1}{p^*_1 + p_1}\right).$$

The bias and MSE of $\bar{y}_{BT}$, to the first order of approximation are, respectively, given by

$$\text{(1.5)} \qquad \operatorname{Bias}(\bar{y}_{BT}) \cong \bar{Y}\left(\frac{1}{n} - \frac{1}{n^*}\right)\left[\frac{1}{2}\rho_{pb_1} C_y C_p - C^2_p\right]$$

and

$$\text{(1.6)} \qquad \operatorname{MSE}(\bar{y}_{BT}) \cong \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{N}\right) C^2_y + \left(\frac{1}{n} - \frac{1}{n^*}\right)\left\{\frac{1}{4}C^2_p - \rho_{pb_1} C_y C_p\right\}\right].$$

2. Proposed two-phase sampling estimator

Following Bahl and Tuteja [2] and Shabbir *et al.* [12], we propose the following exponential difference-cum-ratio type estimator when two auxiliary attributes are available. Also, as mentioned earlier, it is assumed that the population proportion (P_1) and variance ($S^2_{\psi_1}$) for the first auxiliary attribute ψ_1 are unknown, while the same are known for the second auxiliary attribute ψ_2. The proposed estimator is as follows.

$$\begin{aligned}\bar{y}_P = \bar{y} &+ w_1(p_1^* - p_1)\exp\left(\frac{p_1^* - p_1}{p_1^* + p_1}\right) + w_2\left(s^{*2}_{\psi_1} - s^2_{\psi_1}\right)\exp\left(\frac{s^{*2}_{\psi_1} - s^2_{\psi_1}}{s^{*2}_{\psi_1} + s^2_{\psi_1}}\right) \\ &+ w_3(P_2 - p_2^*)\exp\left(\frac{P_2 - p_2^*}{P_2 + p_2^*}\right) + w_4\left(S^2_{\psi_2} - s^{*2}_{\psi_2}\right)\exp\left(\frac{S^2_{\psi_2} - s^{*2}_{\psi_2}}{S^2_{\psi_2} + s^{*2}_{\psi_2}}\right),\end{aligned} \tag{2.1}$$

where w_i, $(i = 1, 2, 3, 4)$ are suitably chosen constants whose values are to be determined.

In the proposed estimator, we try to utilize the information available through the population variances also. There are various studies where this has been done. For details, see, Ahmad *et al.* [1], Dubey and Sharma [3], Jhajj *et al.* [6] and Singh *et al.* [14]. Moreover, in medical and biological studies, it is quite often the case that populations are highly skewed. In such cases, use of information on the variance of the auxiliary attributes can be quite helpful.

Expressing (2.1) in terms of the Δ's, we have

$$\begin{aligned}\bar{y}_P = \bar{Y}(1+\Delta_0) &+ w_1\left\{P_1(1+\Delta^*_{\psi_1}) - P_1(1+\Delta_{\psi_1})\right\} \\ &\qquad \times \exp\left\{\frac{P_1(1+\Delta^*_{\psi_1}) - P_1(1+\Delta_{\psi_1})}{P_1(1+\Delta^*_{\psi_1}) + P_1(1+\Delta_{\psi_1})}\right\} \\ &+ w_2\left\{S^2_{\psi_1}(1+\Delta^*_{\psi_2}) - S^2_{\psi_1}(1+\Delta_{\psi_2})\right\} \\ &\qquad \times \exp\left\{\frac{S^2_{\psi_1}(1+\Delta^*_{\psi_2}) - S^2_{\psi_1}(1+\Delta_{\psi_2})}{S^2_{\psi_1}(1+\Delta^*_{\psi_2}) + S^2_{\psi_1}(1+\Delta_{\psi_2})}\right\} \\ &+ w_3\left\{P_2 - P_2(1+\Delta^*_{\psi_3})\right\}\exp\left\{\frac{P_2 - P_2(1+\Delta^*_{\psi_3})}{P_2 + P_2(1+\Delta^*_{\psi_3})}\right\} \\ &+ w_4\left\{S^2_{\psi_2} - S^2_{\psi_2}(1+\Delta^*_{\psi_4})\right\}\exp\left\{\frac{S^2_{\psi_2} - S^2_{\psi_2}(1+\Delta^*_{\psi_4})}{S^2_{\psi_2} + S^2_{\psi_2}(1+\Delta^*_{\psi_4})}\right\}.\end{aligned} \tag{2.2}$$

Simplifying (2.2) and retaining terms up to order two in the Δ's, we have

$$\begin{aligned}\bar{y}_P - \bar{Y} = \bar{Y}\Delta_0 &+ w_1P_1\left\{(\Delta^*_{\psi_1} - \Delta_{\psi_1}) + \frac{1}{2}(\Delta^*_{\psi_1} - \Delta_{\psi_1})^2\right\} \\ &+ w_2S^2_{\psi_1}\left\{(\Delta^*_{\psi_2} - \Delta_{\psi_2}) + \frac{1}{2}(\Delta^*_{\psi_2} - \Delta_{\psi_2})^2\right\} \\ &- w_3P_2\left\{\Delta^*_{\psi_3} - \frac{1}{2}\Delta^{*2}_{\psi_3}\right\} - w_4S^2_{\psi_2}\left\{\Delta^*_{\psi_4} - \frac{1}{2}\Delta^{*2}_{\psi 4}\right\}.\end{aligned} \tag{2.3}$$

Using (2.3), the bias and MSE of $\bar{y}_P$ to the first order of approximation are, respectively, given by

$$\begin{aligned}\text{Bias}(\bar{y}_P) \cong \frac{1}{2}\Bigg[&\left(\frac{1}{n} - \frac{1}{n^*}\right)\left(w_1P_1C^2_{p_1} + w_2S^2_{\psi_1}(\delta_{040} - 1)\right) \\ &+ \left(\frac{1}{n^*} - \frac{1}{N}\right)\left(w_3P_2C^2_{p_2} + w_4S^2_{\psi_2}(\delta_{004} - 1)\right)\Bigg]\end{aligned} \tag{2.4}$$

and

$$\begin{aligned}\text{MSE}(\bar{y}_P) \cong \left(\frac{1}{n}-\frac{1}{N}\right)\Big\{S_y^2 + w_1^2 S_{\psi_1}^2 + w_2^2 S_{\psi_1}^4(\delta_{040}-1) - 2w_1\rho_{pb_1}S_yS_{\psi_1} \\ - 2w_2S_yS_{\psi_1}^2\delta_{120} + 2w_1w_2S_{\psi_1}^3\delta_{030}\Big\} \\ -\left(\frac{1}{n^*}-\frac{1}{N}\right)\Big\{w_1^2S_{\psi_1}^2 + w_2^2S_{\psi_1}^4(\delta_{040}-1) - w_3^2S_{\psi_2}^2 \\ - w_4^2S_{\psi_2}^4(\delta_{004}-1) - 2w_1\rho_{pb_1}S_yS_{\psi_1} - 2w_2S_yS_{\psi_1}^2\delta_{120} \\ + 2w_3\rho_{pb_2}S_yS_{\psi_2} + 2w_4S_{\psi_2}^2C_y\delta_{102} \\ + 2w_1w_2S_{\psi_1}^3\delta_{030} - 2w_3w_4S_{\psi_2}^3\delta_{003}\Big\}.\end{aligned} \tag{2.5}$$

From (2.5), we get the optimum values of w_i, $(i = 1, 2, 3, 4)$, given by

$$w_1 = \frac{S_y\{(\delta_{040}-1)\rho_{pb_1} - \delta_{030}\delta_{120}\}}{S_{\psi_1}\{(\delta_{040}-1)-\delta_{030}^2\}}, \qquad w_2 = \frac{S_y\{\delta_{120}-\rho_{pb_1}\delta_{030}\}}{S_{\psi_1}^2\{(\delta_{040}-1)-\delta_{030}^2\}},$$
$$w_3 = \frac{S_y\{(\delta_{004}-1)\rho_{pb_2} - \delta_{003}\delta_{102}\}}{S_{\psi_2}\{(\delta_{004}-1)-\delta_{003}^2\}}, \qquad w_4 = \frac{S_y\{\delta_{102}-\rho_{pb_2}\delta_{003}\}}{S_{\psi_2}^2\{(\delta_{004}-1)-\delta_{003}^2\}}.$$

Substituting the optimum values of w_i, $(i = 1, 2, 3, 4)$ in (2.5), we get the minimum MSE of $\bar{y}_P$ in the form:

$$\begin{aligned}\text{MSE}(\bar{y}_P)_{\min} \cong S_y^2\Bigg[\left(\frac{1}{n}-\frac{1}{N}\right) - \left(\frac{1}{n}-\frac{1}{n^*}\right)\left\{\rho_{pb_1}^2 + \frac{(\delta_{120}-\delta_{030}\rho_{pb_1})^2}{\{(\delta_{040}-1)-\delta_{030}^2\}}\right\} \\ - \left(\frac{1}{n^*}-\frac{1}{N}\right)\left\{\rho_{pb_2}^2 + \frac{(\delta_{102}-\delta_{003}\rho_{pb_2})^2}{\{(\delta_{004}-1)-\delta_{003}^2\}}\right\}\Bigg].\end{aligned} \tag{2.6}$$

The expression in (2.6) provides only the ideal minimum MSE of $\bar{y}_P$, since the optimum values of w_i, $(i = 1, 2, 3, 4)$ involve unknown parameters. These values can be replaced by their consistent estimates as discussed by many authors (see Koyuncu and Kadilar [8], Perri and Diana [10] and Pradhan [11]).

3. Comparison of estimators

We now compare the proposed estimator with the other estimators considered here. The following conditions can be verified easily.

Condition (i): $\text{MSE}(\bar{y}_P)_{\min} < \text{Var}(\bar{y})$ if

$$\begin{aligned}S_y^2\Bigg[\left(\frac{1}{n}-\frac{1}{n^*}\right)\left\{\rho_{pb_1}^2 + \frac{(\delta_{120}-\delta_{030}\rho_{pb_1})^2}{\{(\delta_{040}-1)-\delta_{030}^2\}}\right\} \\ + \left(\frac{1}{n^*}-\frac{1}{N}\right)\left\{\rho_{pb_2}^2 + \frac{(\delta_{102}-\delta_{003}\rho_{pb_2})^2}{\{(\delta_{004}-1)-\delta_{003}^2\}}\right\}\Bigg] > 0.\end{aligned}$$

Condition (ii): $\text{MSE}(\bar{y}_P)_{\min} < \text{MSE}(\bar{y}_{BT})$ if

$$\begin{aligned}\bar{Y}^2\Bigg[\left(\frac{1}{n}-\frac{1}{n^*}\right)\left\{\left(\frac{1}{2}C_P - \rho_{pb1}C_y\right)^2 + C_y^2\frac{(\delta_{120}-\delta_{030}\rho_{pb_1})^2}{\{(\delta_{040}-1)-\delta_{030}^2\}}\right\} \\ + \left(\frac{1}{n^*}-\frac{1}{N}\right)\left\{\rho_{pb_2}^2 + \frac{(\delta_{102}-\delta_{003}\rho_{pb_2})^2}{\{(\delta_{004}-1)-\delta_{003}^2\}}\right\}\Bigg] > 0.\end{aligned}$$

Condition (iii): $\mathrm{MSE}\,(\bar{y}_P)_{\min} < \mathrm{MSE}\,(\bar{y}_J)_{\min}$ if

$$S_y^2\left[\left(\frac{1}{n}-\frac{1}{n^*}\right)\left\{\frac{(\delta_{120}-\delta_{030}\rho_{pb_1})^2}{\{(\delta_{040}-1)-\delta_{030}^2\}}\right\} + \left(\frac{1}{n^*}-\frac{1}{N}\right)\left\{\rho_{pb_2}^2+\frac{(\delta_{102}-\delta_{003}\rho_{pb_2})^2}{\{(\delta_{004}-1)-\delta_{003}^2\}}\right\}\right] > 0.$$

The above three conditions will always hold true because $(\delta_{040}-\delta_{030}^2-1) \geq 0$ and $(\delta_{004}-\delta_{003}^2-1) \geq 0$ (see Jhajj et al. [6]).

We use the following numerical examples to further evaluate the performances of various estimators.

Data 1: (Source: Singh and Chaudhary [13], p. 177)

The population consists of 34 wheat farms in 34 villages in certain region of India. The variables are defined as:

y = area under wheat crop (in acres) during 1974,

p_1 = proportion of farms under wheat crop which have more than 500 acres land during 1971, and

p_2 = proportion of farms under wheat crop which have more than 100 acres of land during 1973.

For this data, we have

$$N=34,\quad \bar{Y}=199.4,\quad P_1=0.6765,\quad P_2=0.7353,\quad S_y^2=22564.6,$$
$$S_{\psi_1}^2=0.225490,\quad S_{\psi_2}^2=0.200535,\quad \rho_{pb_1}=0.599,\quad \rho_{pb_2}=0.559,$$
$$\rho_\phi=0.725,\quad \delta_{040}=1.52302,\quad \delta_{004}=2.07490,\quad \delta_{030}=-0.74326,$$
$$\delta_{003}=-1.05086,\quad \delta_{120}=-0.44516,\quad \delta_{102}=-0.58747.$$

Data 2: (Source: Government of Pakistan [4])

The population consists of rice cultivation areas in 73 districts of Pakistan. The variables are defined as:

y = rice production (in 000' tonnes, with one tonne=0.984 ton) during 2003,

p_1 = proportion of farms where rice production is more than 20 tonnes during the year 2002, and

p_2 = proportion of farms with rice cultivation area more than 20 hectares during the year 2003.

For this data, we have

$$N=73,\quad \bar{Y}=61.3,\quad P_1=0.4247,\quad P_2=0.3425,\quad S_y^2=12371.4,$$
$$S_{\psi_1}^2=0.242770,\quad S_{\psi_2}^2=0.228311,\quad \rho_{pb_1}=0.621,\quad \rho_{pb_2}=0.673,$$
$$\rho_\phi=0.889,\quad \delta_{040}=1.16022,\quad \delta_{004}=1.42109,\quad \delta_{030}=0.41703,$$
$$\delta_{003}=0.65939,\quad \delta_{120}=0.25907,\quad \delta_{102}=0.44374.$$

The results for the data sets above are given in Tables 1 and 2.

Table 1. The percentage relative efficiency of estimators with respect to $\bar{y}$ for Data 1

Sample size			Estimator		
n^*	n	$\bar{y}$	$\bar{y}_{BT}$	$\bar{y}_J$	$\bar{y}_P$
10	3	100.00	135.476	138.020	153.384
	7	100.00	114.791	115.680	149.249
15	3	100.00	142.709	145.947	154.589
	7	100.00	129.714	131.748	152.344
	12	100.00	111.785	112.474	148.544
20	3	100.00	146.622	150.262	155.198
	7	100.00	131.731	141.580	153.940
	12	100.00	126.719	128.502	151.772
	17	100.00	111.399	112.062	148.451

Table 2. The percentage relative efficiency of estimators with respect to $\bar{y}$ for Data 2

Sample size			Estimator		
n^*	n	$\bar{y}$	$\bar{y}_{BT}$	$\bar{y}_J$	$\bar{y}_P$
10	3	100.00	127.428	139.182	167.731
	7	100.00	110.845	114.674	175.624
15	3	100.00	132.625	147.435	165.780
	7	100.00	121.056	129.448	170.426
	12	100.00	107.593	110.169	177.564
20	3	100.00	135.386	151.939	164.821
	7	100.00	126.901	138.361	167.941
	12	100.00	116.434	122.639	172.628
	17	100.00	106.118	108.156	178.499
30	3	100.00	138.264	156.728	163.873
	7	100.00	133.339	148.592	165.527
	12	100.00	126.858	138.294	167.958
	20	100.00	119.984	127.851	170.917
	25	100.00	108.077	110.834	177.265

From Tables 1 and 2, one can see that the efficiency of the proposed estimator is greater than the usual mean estimator, the Bahl and Tuetja [2] estimator and the Jhajj *et al.* [5] estimator for both data sets. This was expected based on the efficiency comparisons (Conditions (i-iii)) which always hold true. Also note that the relative efficiencies with respect to the usual mean estimator of all estimators increase as the sample size n^* increases, and also the efficiency of the proposed estimator increases as n increases.

4. Appendix

$$E(\Delta_0) = E(\Delta_{\psi_i}) = E(\Delta^*_{\psi_i}) = 0, (i = 1, 2, 3, 4).$$
$$E(\Delta_0^2) = \left(\frac{1}{n} - \frac{1}{N}\right) C_y^2,\ E(\Delta_{\psi_1}^2) = \left(\frac{1}{n} - \frac{1}{N}\right) C_{p_1}^2,$$
$$E(\Delta_{\psi_2}^2) = \left(\frac{1}{n} - \frac{1}{N}\right) (\delta_{040} - 1),\ E(\Delta_{\psi_3}^2) = \left(\frac{1}{n} - \frac{1}{N}\right) C_{p_2}^2,$$
$$E(\Delta_{\psi_4}^2) = \left(\frac{1}{n} - \frac{1}{N}\right) (\delta_{004} - 1),\ E(\Delta_0\Delta_{\psi_1}) = \left(\frac{1}{n} - \frac{1}{N}\right) \rho_{pb_1} C_y C_{p_1},$$
$$E(\Delta_0\Delta_{\psi_2}) = \left(\frac{1}{n} - \frac{1}{N}\right) C_y \delta_{120},\ E(\Delta_0\Delta_{\psi_3}) = \left(\frac{1}{n} - \frac{1}{N}\right) \rho_{pb_2} C_y C_{p_2},$$
$$E(\Delta_0\Delta_{\psi_4}) = \left(\frac{1}{n} - \frac{1}{N}\right) C_y \delta_{102},\ E(\Delta_{\psi_1}\Delta_{\psi_2}) = \left(\frac{1}{n} - \frac{1}{N}\right) C_{p_1} \delta_{030},$$
$$E(\Delta_{\psi_3}\Delta_{\psi_4}) = \left(\frac{1}{n} - \frac{1}{N}\right) C_{p_2} \delta_{003},\ E(\Delta^{*2}_{\psi_1}) = E(\Delta^*_{\psi_1}\Delta_{\psi_1}) = \left(\frac{1}{n^*} - \frac{1}{N}\right) C_{p_1}^2,$$
$$E(\Delta^{*2}_{\psi_2}) = E(\Delta^*_{\psi_2}\Delta_{\psi_2}) = \left(\frac{1}{n^*} - \frac{1}{N}\right) (\delta_{040} - 1),\ E(\Delta^{*2}_{\psi_3}) = \left(\frac{1}{n^*} - \frac{1}{N}\right) C_{p_2}^2,$$
$$E(\Delta^{*2}_{\psi_4}) = \left(\frac{1}{n^*} - \frac{1}{N}\right) \delta_{004},\ E(\Delta_0\Delta^*_{\psi_1}) = \left(\frac{1}{n^*} - \frac{1}{N}\right) \rho_{pb_1} C_y C_{p_1},$$
$$E(\Delta_0\Delta^*_{\psi_2}) = \left(\frac{1}{n^*} - \frac{1}{N}\right) C_y \delta_{120},\ E(\Delta_0\Delta^*_{\psi_3}) = \left(\frac{1}{n^*} - \frac{1}{N}\right) \rho_{pb_2} C_y C_{p_2},$$
$$E(\Delta_0\Delta^*_{\psi_4}) = \left(\frac{1}{n^*} - \frac{1}{N}\right) C_y \delta_{102},$$
$$E(\Delta^*_{\psi_1}\Delta^*_{\psi_2}) = E(\Delta_{\psi_1}\Delta^*_{\psi_2}) = E(\Delta^*_{\psi_1}\Delta_{\psi_2}) = \left(\frac{1}{n^*} - \frac{1}{N}\right) C_{p_1} \delta_{030},$$
$$E(\Delta^*_{\psi_1}\Delta_{\psi_3}) = E(\Delta_{\psi 1}\Delta^*_{\psi 3}) = E(\Delta^*_{\psi_1}\Delta^*_{\psi_3}) = \left(\frac{1}{n^*} - \frac{1}{N}\right) \rho_\phi C_{p_1} C_{p_2},$$
$$E(\Delta^*_{\psi_1}\Delta^*_{\psi_4}) = E(\Delta_{\psi_1}\Delta^*_{\psi_4}) = E(\Delta^*_{\psi_1}\Delta_{\psi_4}) = \left(\frac{1}{n^*} - \frac{1}{N}\right) \delta_{012} C_{p_1},$$
$$E(\Delta^*_{\psi_2}\Delta^*_{\psi_3}) = E(\Delta_{\psi_2}\Delta^*_{\psi_3}) = E(\Delta^*_{\psi_2}\Delta_{\psi_3}) = \left(\frac{1}{n^*} - \frac{1}{N}\right) \delta_{021} C_{p_2},$$
$$E(\Delta^*_{\psi_2}\Delta^*_{\psi_4}) = E(\Delta_{\psi_2}\Delta^*_{\psi_4}) = E(\Delta^*_{\psi_2}\Delta_{\psi_4}) = \left(\frac{1}{n^*} - \frac{1}{N}\right) (\delta_{022} - 1),$$
$$E(\Delta_{\psi_3}\Delta^*_{\psi_4}) = E(\Delta^*_{\psi_3}\Delta_{\psi_4}) = E(\Delta^*_{\psi_3}\Delta^*_{\psi_4}) = \left(\frac{1}{n^*} - \frac{1}{N}\right) C_{p_2} \delta_{003}.$$

Acknowledgements. We would like to thank the two anonymous referees whose comments helped to improve the presentation of this paper.

References

[1] Ahmad, M. S., Dayyeh, W. A. and Hurairah, A. A. O. *Some estimators for finite population variance under two phase sampling*, Statistics in Transition **6** (1), 143–150, 2003.
[2] Bahl, S. and Tuteja, R. K. *Ratio and product type exponential estimators*, Information and Optimization Sciences **12** (1), 159–163, 1991.

[3] Dubey, V. and Sharma, H. K. *On estimating population variance using auxiliary information*, Statistics in Transition – new series **9** (11), 7–18, 2008.

[4] Government of Pakistan. *Crops Area Production by Districts* (Ministry of Food, Agriculture and Livestock Division, Economic Wing, Pakistan, 2004).

[5] Jhajj, H. S., Sharma, M. K. and Grover, L. K. *A family of estimators of population mean using information on auxiliary attribute*, Pakistan Journal of Statistics **22** (1), 43–50, 2006.

[6] Jhajj, H. S., Sharma, M. K. and Grover, L. K. *An efficient class of chain estimators of population variance under sub-sampling scheme*, Journal of Japan Statistical Society **35** (2), 273–286, 2005.

[7] Kiregyra, B. *Regression type estimators using two auxiliary variables and the models of double sampling from finite populations*, Metrika **31** (3-4), 215–226, 1984.

[8] Koyuncu, N. and Kadilar, C. *Family of estimators of population mean using two auxiliary variables in stratified random sampling*, Communication in Statistics - Theory and Methods **38**, 2398–2417, 2009.

[9] Naik, V. D. and Gupta, P. C. *A note on estimation of mean with known population proportion of an auxiliary character*, Journal of Indian Society of Agricultural Statistics **48** (2), 151–158, 1996.

[10] Perri, P. F. and Diana, G. *Estimating a sensitive proportion through randomized response procedures based on auxiliary information*, Statistical Papers **50**, 661–672, 2009.

[11] Pradhan, B. K. *Some problems of estimation in multiphase sampling* (unpublished PhD thesis, Utkal University, India, 2000).

[12] Shabbir, J., Gupta, S. and Ahmad, S. *Estimation of median in two phase sampling using two auxiliary variables*, Communication in Statistics-Theory and Methods **37** (11), 1815–1822, 2008.

[13] Singh, D. and Chaudhary, F. S. *Theory and Analysis of Sample Survey Designs* (John Wiley and Sons, NewYork, 1986).

[14] Singh, H. P., Upadhyaya, L. N. and Namjoshi, U. D. *Estimation of finite population variance*, Current Science **57** (24), 1331–1334, 1988.

[15] Swain, A. K. P. C. *A note on the use of multiple auxiliary variables in sample surveys*, Trabajos de Estadistica **30**, 135–141, 1970.