Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 70(3) 2016 265–275

Shrinkage Estimator in Dual Frame Survey Sampling

Stephen A. Sedory[1], Sarjinder Singh[1] and Maria del Mar Rueda[2]
[1]*Department of Mathematics, Texas A&M University-Kingsville, Kingsville, TX 78363, USA*
[2]*Department of Statistics and O.R., University of Granada, Avd. Fuentenueva S/N. Granada, Spain*

Received 11 December 2015; Revised 02 September 2016; Accepted 15 September 2016

SUMMARY

In this paper, we have proposed a new shrinkage estimator of the population total for the dual frame survey sampling. The estimator is optimized for three optimizing parameters during the estimation process of the required population total. Thus it is likely that the proposed estimator will be more efficient than the Hartley (1962,1974) estimator listed in Lohr (2011) which makes use of only one optimization. At the end, a similar improvement of Fuller and Burmeister (1972) estimator is suggested.

Keywords: Estimation of total, Dual frame surveys, Efficiency.

1. DUAL FRAME SURVEY BACKGROUND

Following Lohr (2011), we consider a dual frame survey sampling situation as shown in Fig. 1.1 in the absence of auxiliary information. No detail about a review on this topic is given here, because it is available in a recent work of Lohr (2011). Note that in the presence of auxiliary information, it is easy to improve any basic estimator of population mean or total, but it remains more challengeable if an estimator can be improved without using any auxiliary information by following the pioneer ratio estimator of Cochran (1940). Without loss of generality, consider the following situation:

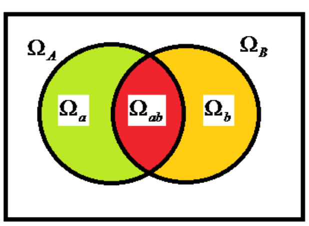

Fig. 1.1 Dual Frame Survey

Let Ω_A be the frame A consisting of N_A number of units in it, and let Ω_B be the frame B consisting of N_B number of units in it. Let $\Omega_a = \Omega_A - \Omega_{ab}$ and $\Omega_b = \Omega_B - \Omega_{ab}$ be the two domains of interest in addition to Ω_{ab} being a common of interest from both frames. Thus, in a dual frame survey, we can write the population total Y as:

$$Y = Y_a + Y_{ab} + Y_b \tag{1.1}$$

where

$$Y_a = \sum_{i\in\Omega_a} y_i,\ Y_{ab} = \sum_{i\in\Omega_{ab}} y_i \text{ and } Y_b = \sum_{i\in\Omega_b} y_i$$

For simplicity, let us consider simple random and without replacement sampling (srswor) so that the inclusion probability of the i^{th} unit to be included in the sample s_A of size n_A from the frame Ω_A is $\pi_i^{(A)} = n_A/N_A$. So the Horvitz and Thompson (1952) estimator of the population total $Y_A = Y_a + Y_{ab}$ of the entire frame Ω_A is given by:

$$\hat{Y}_A = \frac{N_A}{n_A}\sum_{i\in s_A} y_i$$

Corresponding author: Sarjinder Singh
E-mail address: sarjinder@yahoo.com

with variance

$$V(\hat{Y}_A) = \frac{N_A^2(1-f_A)}{n_A}\frac{1}{(N_A-1)}\sum_{i=1}^{N_A}\left(y_i - \bar{Y}_A\right)^2$$

where $f_A = n_A/N_A$ denotes the finite population correction factor in frame Ω_A and $\bar{Y}_A = \frac{1}{N_A}\sum_{i\in\Omega_A} y_i$ is the population mean in the entire frame Ω_A.

Following Haines and Pollock (1998), consider an estimator of the population total Y_a defined as:

$$\hat{Y}_a = \frac{N_A}{n_A}\sum_{i\in s_A}(y_i\delta_i^{(a)}) \tag{1.2}$$

where $\delta_i^{(a)} = \begin{cases}1 & \text{if} \quad i\in\Omega_a \\ 0 & \text{otherwise}\end{cases}$ (1.3)

No doubt

$$E(\hat{Y}_a) = \sum_{i\in\Omega_A}\left(y_i\delta_i^{(a)}\right)$$

$$= \sum_{i\in\Omega_a} y_i\delta_i^{(a)} + \sum_{i\in\Omega_{ab}} y_i\delta_i^{(a)} = Y_a + 0 = Y_a \tag{1.4}$$

It is true that the total will not be affected by changing values with zero, but variance will get effected.

Now the question is: What is the variance of $\hat{Y}_a$?

Please recall that the original $y_i\delta_i^{(a)}$ has first order inclusion probability $\pi_i^{(A)} = \frac{n_A}{N_A}$, $\left(\text{not with } \pi_i^{(a)} = \frac{n_a}{N_a}\right)$ and the pair $\left(y_i\delta_i^{(a)}\right.$ and $\left.y_j\delta_j^{(a)}\right)$ has second order inclusion probability $\pi_{ij}^{(A)} = \frac{n_A(n_A-1)}{N_A(N_A-1)}\left(\text{not with } \pi_{ij}^{(a)} = \frac{n_a(n_a-1)}{N_a(N_a-1)}\right)$.

Then the variance of $\hat{Y}_a$ will be:

$$V(\hat{Y}_a) = \frac{N_A^2(1-f_A)}{n_A}\frac{1}{(N_A-1)}\sum_{i\in\Omega_A}\left(y_i\delta_i^{(a)} - \frac{1}{N_A}\sum_{i\in\Omega_A}\delta_i^{(a)}y_i\right)^2$$

$$= \frac{N_A^2(1-f_A)}{n_A}\frac{1}{(N_A-1)}\sum_{i\in\Omega_A}\left(y_i\delta_i^{(a)} - \frac{1}{N_A}\sum_{i\in\Omega_a}y_i\right)^2$$

$$= \frac{N_A^2(1-f_A)}{n_A}\frac{1}{(N_A-1)}\sum_{i\in\Omega_A}\left(y_i\delta_i^{(a)} - \frac{N_a}{N_A}\bar{Y}_a\right)^2$$

$$= \frac{N_A^2(1-f_A)}{n_A}\frac{1}{(N_A-1)}$$

$$\left[\sum_{i\in\Omega_a}\left(y_i - \frac{N_a}{N_A}\bar{Y}_a\right)^2 + \sum_{i\in\Omega_{ab}}\left(0 - \frac{N_a}{N_A}\bar{Y}_a\right)^2\right]$$

$$= \frac{N_A^2(1-f_A)}{n_A}\frac{1}{(N_A-1)}$$

$$\left[\sum_{i\in\Omega_a}\left(y_i - \bar{Y}_a + \bar{Y}_a - \frac{N_a}{N_A}\bar{Y}_a\right)^2 + \sum_{i\in\Omega_{ab}}\left(0 - \frac{N_a}{N_A}\bar{Y}_a\right)^2\right]$$

$$= \frac{N_A^2(1-f_A)}{n_A}\frac{1}{(N_A-1)}$$

$$\left[\sum_{i\in\Omega_a}\left(y_i - \bar{Y}_a\right)^2 + N_a\bar{Y}_a^2\left(1 - \frac{N_a}{N_A}\right)^2 + N_{ab}\left(\frac{N_a}{N_A}\bar{Y}_a\right)^2\right] \tag{1.5}$$

In the same way, defining

$$\hat{Y}_{ab} = \frac{N_A}{n_A}\sum_{i\in s_A}\left(y_i\delta_i^{(ab)}\right) \tag{1.6}$$

where

$$\delta_i^{(ab)} = \begin{cases}1 & \text{if} \quad i\in\Omega_{ab} \\ 0 & \text{otherwise}\end{cases} \tag{1.7}$$

We have

$$E(\hat{Y}_{ab}) = E\left[\frac{N_A}{n_A}\sum_{i\in s_A}\left(y_i\delta_i^{(ab)}\right)\right] = Y_{ab} \tag{1.8}$$

and

$$V(\hat{Y}_{ab}) = V\left[\frac{N_A}{n_A}\sum_{i\in s_A}(y_i\delta_i^{(ab)})\right] = \frac{N_A^2(1-f_A)}{n_A}\frac{1}{(N_A-1)}$$

$$\left[\sum_{i\in\Omega_{ab}}\left(y_i - \bar{Y}_{ab}\right)^2 + N_{ab}\bar{Y}_{ab}^2\left(1 - \frac{N_{ab}}{N_A}\right)^2 + N_a\left(\frac{N_{ab}}{N_A}\bar{Y}_{ab}\right)^2\right] \tag{1.9}$$

Note that the variance of the dual frame estimator of the total Y_A is given by:

$$V\left(\hat{Y}_A^{(dual)}\right) = V\left(\hat{Y}_a + \hat{Y}_{ab}\right)$$

$$= V(\hat{Y}_a) + V(\hat{Y}_{ab}) + 2Cov(\hat{Y}_a, \hat{Y}_{ab}) \tag{1.10}$$

Now we show that the estimators $\hat{Y}_a$ and $\hat{Y}_{ab}$ in (1.10) are always dependent and given by:

$$Cov\left(\hat{Y}_a, \hat{Y}_{ab}\right) = Cov\left[\frac{N_A}{n_A}\sum_{i\in s_A} y_i\delta_i^{(a)}, \frac{N_A}{n_A}\sum_{i\in s_A} y_i\delta_i^{(ab)}\right]$$

$$= \frac{N_A^2(1-f_A)}{n_A(N_A-1)}\left[\sum_{i\in\Omega_A} y_i^2\delta_i^{(a)}\delta_i^{(ab)} - \frac{\left(\sum_{i\in\Omega_A} y_i\delta_i^{(a)}\right)\left(\sum_{i\in\Omega_A} y_i\delta_i^{(ab)}\right)}{N_A}\right]$$

$$= \frac{N_A^2(1-f_A)}{n_A(N_A-1)}\left[0 - \frac{\left(\sum_{i\in\Omega_a} y_i\right)\left(\sum_{i\in\Omega_{ab}} y_i\right)}{N_A}\right]$$

$$= \frac{N_A^2(1-f_A)}{n_A(N_A-1)}\left[0 - \frac{N_a N_{ab}\left(\sum_{i\in\Omega_a} y_i\right)\left(\sum_{i\in\Omega_{ab}} y_i\right)}{N_A N_a N_{ab}}\right]$$

$$= -\frac{N_A^2(1-f_A)N_a N_{ab}}{n_A N_A(N_A-1)}\bar{Y}_a\bar{Y}_{ab}.$$

The variance of the estimator $\hat{Y}_A^{(dual)}$ is given by:

$$V\left(\hat{Y}_A^{(dual)}\right) = \frac{N_A^2(1-f_A)}{n_A}\frac{1}{(N_A-1)}$$

$$\left[\sum_{i\in\Omega_a}\left(y_i-\bar{Y}_a\right)^2 + N_a\bar{Y}_a^2\left(1-\frac{N_a}{N_A}\right)^2 + N_{ab}\left(\frac{N_a}{N_A}\bar{Y}_a\right)^2\right]$$

$$+\frac{N_A^2(1-f_A)}{n_A}\frac{1}{(N_A-1)}$$

$$\left[\sum_{i\in\Omega_{ab}}\left(y_i-\bar{Y}_{ab}\right)^2 + N_{ab}\bar{Y}_{ab}^2\left(1-\frac{N_{ab}}{N_A}\right)^2 + N_a\left(\frac{N_{ab}}{N_A}\bar{Y}_{ab}\right)^2\right]$$

$$-2\frac{N_A^2(1-f_A)N_a N_{ab}}{n_A N_A(N_A-1)}\bar{Y}_a\bar{Y}_{ab}$$

$$= \frac{N_A^2(1-f_A)}{n_A(N_A-1)}\left[\sum_{i\in\Omega_a}(y_i-\bar{Y}_a)^2 + \sum_{i\in\Omega_{ab}}(y_i-\bar{Y}_{ab})^2\right.$$

$$+N_a\bar{Y}_a^2\left(1-\frac{N_a}{N_A}\right)^2 + N_{ab}\left(\frac{N_a}{N_A}\bar{Y}_a\right)^2$$

$$\left.+N_{ab}\bar{Y}_{ab}^2\left(1-\frac{N_{ab}}{N_A}\right)^2 + N_a\left(\frac{N_{ab}}{N_A}\bar{Y}_{ab}\right)^2 - 2\frac{N_a N_{ab}}{N_A}\bar{Y}_a\bar{Y}_{ab}\right]$$

$$= \frac{N_A^2(1-f_A)}{n_A(N_A-1)}\left[(N_A-1)S_A^2 - N_a(\bar{Y}_a-\bar{Y}_A)^2\right.$$

$$-N_{ab}(\bar{Y}_{ab}-\bar{Y}_A)^2 + N_a\bar{Y}_a^2\left(1-\frac{N_a}{N_A}\right)^2 + N_{ab}\left(\frac{N_a}{N_A}\bar{Y}_a\right)^2$$

$$\left.+N_{ab}\bar{Y}_{ab}^2\left(1-\frac{N_{ab}}{N_A}\right)^2 + N_a\left(\frac{N_{ab}}{N_A}\bar{Y}_{ab}\right)^2 - 2\frac{N_a N_{ab}}{N_A}\bar{Y}_a\bar{Y}_{ab}\right]$$

$$= V\left(\hat{Y}_A\right) + \frac{N_A^2(1-f_A)}{n_A(N_A-1)}\left[N_a\bar{Y}_a^2\left(1-\frac{N_a}{N_A}\right)^2 + N_{ab}\left(\frac{N_a}{N_A}\bar{Y}_a\right)^2\right.$$

$$+N_{ab}\bar{Y}_{ab}^2\left(1-\frac{N_{ab}}{N_A}\right)^2 + N_a\left(\frac{N_{ab}}{N_A}\bar{Y}_{ab}\right)^2$$

$$\left.-N_a(\bar{Y}_a-\bar{Y}_A)^2 - N_{ab}(\bar{Y}_{ab}-\bar{Y}_A)^2 - 2\frac{N_a N_{ab}}{N_A}\bar{Y}_a\bar{Y}_{ab}\right]$$

$$= V\left(\hat{Y}_A\right) + \frac{N_A^2(1-f_A)}{n_A(N_A-1)}\left[N_a\bar{Y}_a^2\left(1-\frac{N_a}{N_A}\right)^2 + N_{ab}\left(\frac{N_a}{N_A}\bar{Y}_a\right)^2\right.$$

$$+N_{ab}\bar{Y}_{ab}^2\left(1-\frac{N_{ab}}{N_A}\right)^2 + N_a\left(\frac{N_{ab}}{N_A}\bar{Y}_{ab}\right)^2 - N_a\left(\bar{Y}_a^2+\bar{Y}_A^2-2\bar{Y}_a\bar{Y}_A\right)$$

$$\left.-N_{ab}\left(\bar{Y}_{ab}^2+\bar{Y}_A^2-2\bar{Y}_{ab}\bar{Y}_A\right) - 2\frac{N_a N_{ab}}{N_A}\bar{Y}_a\bar{Y}_{ab}\right]$$

$$= V\left(\hat{Y}_A\right) + \frac{N_A^2(1-f_A)}{n_A(N_A-1)}\left[N_a\bar{Y}_a^2\left(1-\frac{N_a}{N_A}\right)^2 + N_{ab}\left(\frac{N_a}{N_A}\bar{Y}_a\right)^2\right.$$

$$+N_{ab}\bar{Y}_{ab}^2\left(1-\frac{N_{ab}}{N_A}\right)^2 + N_a\left(\frac{N_{ab}}{N_A}\bar{Y}_{ab}\right)^2 - N_a\bar{Y}_a^2 - N_a\bar{Y}_A^2$$

$$\left.+2N_a\bar{Y}_a\bar{Y}_A - N_{ab}\bar{Y}_{ab}^2 - N_{ab}\bar{Y}_A^2 + 2N_{ab}\bar{Y}_{ab}\bar{Y}_A - 2\frac{N_a N_{ab}}{N_A}\bar{Y}_a\bar{Y}_{ab}\right]$$

$$= V\left(\hat{Y}_A\right) + \frac{N_A^2(1-f_A)}{n_A(N_A-1)}\left[N_a\bar{Y}_a^2\left(1-\frac{N_a}{N_A}\right)^2 + N_{ab}\left(\frac{N_a}{N_A}\bar{Y}_a\right)^2\right.$$

$$+N_{ab}\bar{Y}_{ab}^2\left(1-\frac{N_{ab}}{N_A}\right)^2 + N_a\left(\frac{N_{ab}}{N_A}\bar{Y}_{ab}\right)^2 - N_a\bar{Y}_a^2 - N_a\bar{Y}_A^2$$

$$\left.+2N_a\bar{Y}_a\bar{Y}_A - N_{ab}\bar{Y}_{ab}^2 - N_{ab}\bar{Y}_A^2 + 2N_{ab}\bar{Y}_{ab}\bar{Y}_A - 2\frac{N_a N_{ab}}{N_A}\bar{Y}_a\bar{Y}_{ab}\right] \quad (1.11)$$

Now note that:

$$\Bigg[N_a\bar{Y}_a^2\left(1-\frac{N_a}{N_A}\right)^2+N_{ab}\left(\frac{N_a}{N_A}\bar{Y}_a\right)^2+N_{ab}\bar{Y}_{ab}^2\left(1-\frac{N_{ab}}{N_A}\right)^2$$

$$+N_a\left(\frac{N_{ab}}{N_A}\bar{Y}_{ab}\right)^2-N_a\bar{Y}_a^2-N_a\bar{Y}_A^2+2N_a\bar{Y}_a\bar{Y}_A-N_{ab}\bar{Y}_{ab}^2$$

$$-N_{ab}\bar{Y}_A^2+2N_{ab}\bar{Y}_{ab}\bar{Y}_A-2\frac{N_aN_{ab}}{N_A}\bar{Y}_a\bar{Y}_{ab}\Bigg]$$

$$=N_a^2\bar{Y}_a^2+\frac{N_a^3\bar{Y}_a^2}{N_A^2}-2\frac{\bar{Y}_a^2N_a^2}{N_A}+\frac{N_{ab}N_a^2}{N_A^2}\bar{Y}_a^2+N_{ab}\bar{Y}_{ab}^2$$

$$+\frac{N_{ab}^3}{N_A^2}\bar{Y}_{ab}^2-2\frac{N_{ab}^2}{N_A}\bar{Y}_{ab}^2+\frac{N_aN_{ab}^2\bar{Y}_{ab}^2}{N_A^2}-N_a\bar{Y}_a^2$$

$$-N_a\bar{Y}_A^2+2N_a\bar{Y}_a\bar{Y}_A-N_{ab}\bar{Y}_{ab}^2-N_{ab}\bar{Y}_A^2$$

$$+2N_{ab}\bar{Y}_{ab}\bar{Y}_A-2\frac{N_aN_{ab}\bar{Y}_a\bar{Y}_{ab}}{N_A}=0 \qquad (1.12)$$

From (1.11) and (1.12)

$$V\left(\hat{Y}_A^{(dual)}\right)=\frac{N_A^2\left(1-f_A\right)}{n_A}S_A^2 \qquad (1.13)$$

where

$$S_A^2=\frac{1}{N_A-1}\sum_{i\in\Omega_A}\left(y_i-\bar{Y}_A\right)^2.$$

Hence the variance of the estimator of population total using additional information remains the same. Note that these derivations derived in different parts may be useful in another research work by researchers although immediately nothing is coming to mind.

Now let us consider the problem of estimation of the population total $Y = Y_a + Y_{ab} + Y_b$ with the estimator considered in Lohr (2011) defined as:

$$\hat{Y}_{dual}=\hat{Y}_a+\alpha\hat{Y}_{ab}+(1-\alpha)\hat{Y}_{ba}+\hat{Y}_b \qquad (1.14)$$

or

$$\hat{Y}_{dual}=\hat{Y}_a+\hat{Y}_{ab}+\left(\alpha-1\right)\hat{Y}_{ab}-\alpha\hat{Y}_{ba}+\hat{Y}_{ba}+\hat{Y}_b$$

or

$$\hat{Y}_{dual}=\hat{Y}_A+\left(\alpha-1\right)\hat{Y}_{ab}-\alpha\hat{Y}_{ba}+\hat{Y}_B$$

By the definition of variance, we have:

$$V\left(\hat{Y}_{dual}\right)=V(\hat{Y}_A)+V(\hat{Y}_B)+(\alpha-1)^2V\left(\hat{Y}_{ab}\right)$$

$$+\alpha^2V\left(\hat{Y}_{ba}\right)+2\left(\alpha-1\right)Cov\left(\hat{Y}_A,\hat{Y}_{ab}\right)-2\alpha Cov\left(\hat{Y}_{ba},\hat{Y}_B\right)$$

$$=V(\hat{Y}_A)+V(\hat{Y}_B)+V\left(\hat{Y}_{ab}\right)-2Cov\left(\hat{Y}_A,\hat{Y}_{ab}\right)$$

$$+\alpha^2\left[V\left(\hat{Y}_{ab}\right)+V\left(\hat{Y}_{ba}\right)\right]-2\alpha V\left(\hat{Y}_{ab}\right)$$

$$+2\alpha Cov\left(\hat{Y}_A,\hat{Y}_{ab}\right)-2\alpha Cov\left(\hat{Y}_{ba},\hat{Y}_B\right) \qquad (1.15)$$

The optimum value which minimizes the variance of the estimator in (1.14) is given by:

$$\alpha=\frac{V\left(\hat{Y}_{ab}\right)+Cov\left(\hat{Y}_{ba},\hat{Y}_B\right)-Cov\left(\hat{Y}_A,\hat{Y}_{ab}\right)}{V(\hat{Y}_{ab})+V(\hat{Y}_{ba})} \qquad (1.16)$$

Note that this optimum value of α is same as reported by Lohr (2011, Survey Methodology, page 200) because:

$$Cov\left(\hat{Y}_A,\hat{Y}_{ab}\right)=Cov\left(\hat{Y}_a+Y_{ab},\hat{Y}_{ab}\right)$$

$$=Cov\left(\hat{Y}_a,\hat{Y}_{ab}\right)+Cov\left(\hat{Y}_{ab},\hat{Y}_{ab}\right)$$

$$=Cov\left(\hat{Y}_a,\hat{Y}_{ab}\right)+V\left(\hat{Y}_{ab}\right)$$

and

$$Cov\left(\hat{Y}_B,\hat{Y}_{ba}\right)=Cov\left(\hat{Y}_b+Y_{ba},\hat{Y}_{ba}\right)$$

$$=Cov\left(\hat{Y}_b,\hat{Y}_{ba}\right)+V\left(\hat{Y}_{ba}\right)$$

The minimum variance in dual frame survey sampling is given by:

$$Min\,V(\hat{Y}_{dual})=V(\hat{Y}_A)+V(\hat{Y}_B)+V\left(\hat{Y}_{ab}\right)$$

$$-2\,Cov\left(\hat{Y}_A,\hat{Y}_{ab}\right)$$

$$-\frac{\left[V(\hat{Y}_{ab})-Cov\left(\hat{Y}_A,\hat{Y}_{ab}\right)+Cov\left(\hat{Y}_B,\hat{Y}_{ba}\right)\right]^2}{V(\hat{Y}_{ab})+V(\hat{Y}_{ba})} \qquad (1.17)$$

where

$$V\left(\hat{Y}_B^{(dual)}\right)=\frac{N_B^2(1-f_B)}{n_B}S_B^2 \qquad (1.18)$$

where

$$S_B^2 = \frac{1}{N_B - 1} \sum_{i \in \Omega_B} \left(y_i - \bar{Y}_B\right)^2.$$

$$Cov\left(\hat{Y}_a, \hat{Y}_{ab}\right) = -\frac{N_A^2\left(1 - f_A\right) N_a N_{ab}}{n_A N_A \left(N_A - 1\right)} \bar{Y}_a \bar{Y}_{ab},$$

and

$$Cov\left(\hat{Y}_b, \hat{Y}_{ba}\right) = -\frac{N_B^2(1 - f_B) N_b N_{ab}}{n_B N_B (N_B - 1)} \bar{Y}_b \bar{Y}_{ab}$$

In the next section, we define a few notations which remain useful in deriving the bias and mean squared error of the proposed shrinkage estimator.

2. NOTATIONS

Let us define

$$\in_a = \frac{\hat{Y}_a}{Y_a} - 1, \in_b = \frac{\hat{Y}_b}{Y_b} - 1, \in_{ab} = \frac{\hat{Y}_{ab}}{Y_{ab}} - 1 \text{ and}$$

$$\in_{ba} = \frac{\hat{Y}_{ba}}{Y_{ab}} - 1$$

Such that

$$E\left(\in_a\right) = E\left(\in_b\right) = E\left(\in_{ba}\right) = 0$$

and

$$E\left(\in_a^2\right) = V\left(\hat{Y}_a\right) \big/ Y_a^2,\ E\left(\in_b^2\right) = V\left(\hat{Y}_b\right) \big/ Y_b^2,$$

$$E\left(\in_{ab}^2\right) = V\left(\hat{Y}_{ab}\right) \big/ Y_{ab}^2,\ E\left(\in_{ba}^2\right) = V\left(\hat{Y}_{ba}\right) \big/ Y_{ab}^2,$$

$$E\left(\in_a \in_b\right) = 0,\ E\left(\in_a \in_{ab}\right) = Cov(\hat{Y}_a, \hat{Y}_{ab}) \big/ (Y_a Y_{ab}),$$

$$E\left(\in_a \in_{ba}\right) = 0,\ E\left(\in_b \in_{ab}\right) = 0,$$

$$E\left(\in_b \in_{ba}\right) = Cov(\hat{Y}_b, \hat{Y}_{ba}) \big/ (Y_b Y_{ab})$$

$$\text{and } E\left(\in_{ab} \in_{ba}\right) = 0.$$

Searls (1964) was the first to suggest a shrinkage estimator of the population mean, and in the next section we also suggest a new shrinkage estimator of the population total in dual frame surveys. The Searls' estimator is found to be efficient in case of small samples, and consistent for large sample cases. Thus the motivation and demand of construction of new shrinkage estimators is well known in the literature of survey sampling. It also motivates to investigate the behavior of a shrinkage estimator in case of dual frame surveys without making use of any auxiliary information. It is a fact that many dual-frame surveys continue to rely on expensive data collection and past experience protocols, thus shrinkage estimator may help such surveys.

3. PROPOSED SHRINKAGE ESTIMATOR

We suggest a new shrinkage estimator for a dual frame survey sampling defined as:

$$\hat{Y}_{\text{Shrink}} = \lambda_1 \hat{Y}_a + \lambda_2 \hat{Y}_b + \lambda_3 \hat{Y}_{ab} + (1 - \lambda_3) \hat{Y}_{ba} \quad (3.1)$$

where λ_1, λ_2 and λ_3 are constants to be determined such that the mean squared error of the proposed estimator is minimum. If $\lambda_1 = \lambda_2 = 1$, then the proposed estimator reduces to the pioneer estimator of Hartley (1962, 1974).

The estimator (3.1) in terms of $\in_a$, $\in_b$, $\in_{ab}$ and $\in_{ba}$ can be written as:

$$\hat{Y}_{\text{Shrink}} = Y + (\lambda_1 - 1)Y_a + (\lambda_2 - 1)Y_b + \lambda_1 Y_a \in_a + \lambda_2 Y_b \in_b$$
$$+ \lambda_3 Y_{ab} \in_{ab} + Y_{ab}(1 - \lambda_3) \in_{ba} \quad (3.2)$$

Taking expected value on both sides, and using results from section 2, we get the bias in the proposed shrinkage estimator as:

$$B\left(\hat{Y}_{\text{Shrink}}\right) = E\left(\hat{Y}_{\text{Shrink}}\right) - Y$$
$$= \left(\lambda_1 - 1\right) Y_a + \left(\lambda_2 - 1\right) Y_b \quad (3.3)$$

The mean squared error of the proposed shrinkage estimator $\hat{Y}_{\text{Shrink}}$ is given by:

$$MSE\left(\hat{Y}_{\text{Shrink}}\right) = E\left[\hat{Y}_{\text{Shrink}} - Y\right]^2$$
$$= E\Big[\left(\lambda_1 - 1\right) Y_a + \left(\lambda_2 - 1\right) Y_b + \lambda_1 Y_a \in_a + \lambda_2 Y_b \in_b$$
$$+ \lambda_3 Y_{ab} \in_{ab} + Y_{ab} \left(1 - \lambda_3\right) \in_{ba}\Big]^2$$
$$= \left(\lambda_1 - 1\right)^2 Y_a^2 + \left(\lambda_2 - 1\right)^2 Y_b^2 + 2\left(\lambda_1 - 1\right)\left(\lambda_2 - 1\right) Y_a Y_b$$
$$+ \lambda_1^2 V\left(\hat{Y}_a\right) + \lambda_2^2 V\left(\hat{Y}_b\right) + \lambda_3^2 V\left(\hat{Y}_{ab}\right)$$
$$+ \left(1 - \lambda_3\right)^2 V\left(\hat{Y}_{ba}\right) + 2\lambda_1 \lambda_3 Cov\left(\hat{Y}_a, \hat{Y}_{ab}\right)$$
$$+ 2\lambda_2 \left(1 - \lambda_3\right) Cov\left(\hat{Y}_b, \hat{Y}_{ba}\right) \quad (3.4)$$

The mean squared error in (3.3) will be minimum for the optimum values of λ_1, λ_2 and λ_3 obtained by solving a set of three linear equations as:

$$\lambda_1\left[Y_a^2+V\left(\hat{Y}_a\right)\right]+\lambda_2 Y_a Y_b+\lambda_3 Cov\left(\hat{Y}_a,\hat{Y}_{ab}\right)$$

$$=Y_a^2+Y_aY_b \tag{3.5}$$

$$\lambda_1 Y_a Y_b+\lambda_2\left[Y_b^2+V(\hat{Y}_b)\right]-\lambda_3 Cov(\hat{Y}_b,\hat{Y}_{ba})$$

$$=Y_b^2+Y_aY_b-Cov\left(\hat{Y}_b,\hat{Y}_{ba}\right) \tag{3.6}$$

and

$$\lambda_1 Cov(\hat{Y}_a,\hat{Y}_{ab})-\lambda_2 Cov\left(\hat{Y}_b,\hat{Y}_{ba}\right)$$

$$+\lambda_3\left[V(\hat{Y}_{ab})+V(\hat{Y}_{ba})\right]=V\left(\hat{Y}_{ba}\right) \tag{3.7}$$

The optimum values of λ_1, λ_2 and λ_3 are given by:

$$\lambda_1=\frac{CD-BE}{AD-B^2} \tag{3.8}$$

$$\lambda_2=\frac{AE-BC}{AD-B^2} \tag{3.9}$$

and

$$\lambda_3=\frac{[V(\hat{Y}_{ba})(AD-B^2)-(CD-BE)Cov(\hat{Y}_a,\hat{Y}_{ab})+(AE-BC)Cov(\hat{Y}_b,\hat{Y}_{ba})]}{(AD-B^2)[V(\hat{Y}_{ab})+V(\hat{Y}_{ba})]} \tag{3.10}$$

where

$$A=\left(Y_a^2+V(\hat{Y}_a)\right)\left(V(\hat{Y}_{ab})+V(\hat{Y}_{ba})\right)-Cov^2\left(\hat{Y}_a,\hat{Y}_{ab}\right) \tag{3.11}$$

$$B=Y_aY_b\left[V\left(\hat{Y}_{ab}\right)+V\left(\hat{Y}_{ba}\right)\right]+Cov\left(\hat{Y}_a,\hat{Y}_{ab}\right)Cov\left(\hat{Y}_b,\hat{Y}_{ba}\right) \tag{3.12}$$

$$C=\left[V\left(\hat{Y}_{ab}\right)+V\left(\hat{Y}_{ba}\right)\right]\left[Y_a^2+Y_aY_b\right]-Cov\left(\hat{Y}_a,\hat{Y}_{ab}\right)V\left(\hat{Y}_{ba}\right) \tag{3.13}$$

$$D=\left[V\left(\hat{Y}_{ab}\right)+V\left(\hat{Y}_{ba}\right)\right]\left[Y_b^2+V\left(\hat{Y}_b\right)\right]-Cov^2\left(\hat{Y}_b,\hat{Y}_{ba}\right) \tag{3.14}$$

and

$$E=\left[V\left(\hat{Y}_{ab}\right)+V\left(\hat{Y}_{ba}\right)\right]\left[Y_b^2+Y_aY_b-Cov\left(\hat{Y}_b,\hat{Y}_{ba}\right)\right]+V\left(\hat{Y}_{ba}\right)Cov\left(\hat{Y}_b,\hat{Y}_{ba}\right) \tag{3.15}$$

The resultant minimum mean squared error of the proposed shrinkage estimator is given by:

$$\begin{aligned}\text{Min.MSE}\left(\hat{Y}_{\text{shrink}}\right)=\frac{1}{\left(AD-B^2\right)^2}\Bigg[&\left\{\left(D(C-A)-B(E-B)\right)Y_a+\left(A(E-D)-B(C-B)\right)Y_b\right\}^2\\&+\left(CD-BE\right)^2V\left(\hat{Y}_a\right)+\left(AE-BC\right)^2V\left(\hat{Y}_b\right)\\&+\frac{\left[V(\hat{Y}_{ab})\left\{V(\hat{Y}_{ba})(AD-B^2)-(CD-BE)Cov(\hat{Y}_a,\hat{Y}_{ab})+(AE-BC)Cov(\hat{Y}_b,\hat{Y}_{ba})\right\}^2\right]}{[V(\hat{Y}_{ab})+V(\hat{Y}_{ba})]^2}\\&\frac{\left[V(\hat{Y}_{ba})\left\{\left(AD-B^2\right)V\left(\hat{Y}_{ab}\right)+\left(CD-BE\right)Cov\left(\hat{Y}_a,\hat{Y}_{ab}\right)+\left(BC-AE\right)Cov\left(\hat{Y}_b,\hat{Y}_{ba}\right)\right\}^2\right]}{\left[V\left(\hat{Y}_{ab}\right)+V\left(\hat{Y}_{ba}\right)\right]^2}\\&+2\frac{\left[Cov\left(\hat{Y}_a,\hat{Y}_{ab}\right)(CD-BE)\{V\left(\hat{Y}_{ba}\right)\left(AD-B^2\right)-(CD-BE)Cov\left(\hat{Y}_a,\hat{Y}_{ab}\right)+(AE-BC)Cov\left(\hat{Y}_b,\hat{Y}_{ba}\right)\}\right]}{[V(\hat{Y}_{ab})+V(\hat{Y}_{ba})]}\\&+2\frac{\left[Cov\left(\hat{Y}_b,\hat{Y}_{ba}\right)(AE-BC)\left\{V\left(\hat{Y}_{ab}\right)\left(AD-B^2\right)+(CD-BE)\,Cov\left(\hat{Y}_a,\hat{Y}_{ab}\right)+(BC-AE)Cov\left(\hat{Y}_b,\hat{Y}_{ba}\right)\right\}\right]}{[V(\hat{Y}_{ab})+V(\hat{Y}_{ba})]}\Bigg]\end{aligned} \tag{3.16}$$

The proposed shrinkage estimator is not easily comparable with the Hartley (1962, 1974) estimator, thus we perform a simulation study in the next section. In the next section, we show that the proposed shrinkage estimator performs very well in comparison to the pioneer Hartley (1962, 1974) estimator.

4. SIMULATION STUDY

We define exact percent relative bias in the proposed shrinkage estimator as:

$$\mathrm{RB}\left(\hat{Y}_{\mathrm{Shrink}}\right) = \frac{(\lambda_1 - 1) Y_a + (\lambda_2 - 1) Y_b}{Y} \times 100\% \tag{4.1}$$

and, the percent relative efficiency of the proposed shrinkage estimator $\hat{Y}_{\mathrm{Shrink}}$ with respect to the Hartley (1962,1974) estimator as:

$$\mathrm{RE} = \frac{V\left(\hat{Y}_{\mathrm{dual}}\right)}{\mathrm{MSE}\left(\hat{Y}_{\mathrm{Shrink}}\right)} \times 100\% \tag{4.2}$$

We consider several situations in the simulation study to investigate whether the proposed shrinkage estimator performs better than the pioneer Hartley's estimator. We assume Y_a = 2000, Y_b = 2500, and then we consider four different values of Y_a = 500, 1000, 1500 and 2000. These four choices of values of the overlap totals can be considered as small, medium, large and complete overlap of two frames with each other. We also considered three different values of $V\left(\hat{Y}_a\right)$ = 50, 100 and 150; three different values of $V\left(\hat{Y}_b\right)$ = 50, 100 and 150; three different values of $V\left(\hat{Y}_{ab}\right)$ = 50, 100 and 150; and three different values of $V\left(\hat{Y}_{ba}\right)$ = 50, 100 and 150. Note that the values of $Cov\left(\hat{Y}_a, \hat{Y}_{ab}\right)$ and $Cov\left(\hat{Y}_b, \hat{Y}_{ba}\right)$ are always negative values. Thus we consider different nine negative values of the correlation coefficient Corr $\left(\hat{Y}_a, \hat{Y}_{ab}\right)$ between –0.9 to –0.1 with an increasing step of 0.1; and also we consider different nine negative values of the correlation coefficient Corr $\left(\hat{Y}_b, \hat{Y}_{ba}\right)$ between –0.9 to –0.1 with an increasing step of 0.1. From the simulation study we found that the choice of values of Y_{ab}, $V\left(\hat{Y}_a\right)$, $V\left(\hat{Y}_b\right)$, $V\left(\hat{Y}_{ab}\right)$ and $V\left(\hat{Y}_{ba}\right)$ are less important than a choice of values of correlation coefficients $\mathrm{Corr}\left(\hat{Y}_a, \hat{Y}_{ab}\right)$ and $\mathrm{Corr}\left(\hat{Y}_b, \hat{Y}_{ba}\right)$. The values of the correlation coefficients $\mathrm{Corr}\left(\hat{Y}_a, \hat{Y}_{ab}\right)$ and $\mathrm{Corr}\left(\hat{Y}_b, \hat{Y}_{ba}\right)$ are in fact used to compute the values of $\mathrm{Corr}\left(\hat{Y}_a, \hat{Y}_{ab}\right)$ and $\mathrm{Corr}\left(\hat{Y}_b, \hat{Y}_{ba}\right)$. We provide the FORTRAN Codes used in the simulation study in Appendix I.

Table 4.1 is devoted to present the average value of percent relative efficiency (RE) of the proposed shrinkage estimator along with its standard deviation in the simulation study for different choice of overlaps considered as small, medium, large and complete overlaps. It also gives minimum, median and maximum relative efficiency values. There are 1965 cases where the proposed estimator shows a value of percent relative efficiency between 101.01% and 430.71% with a median value of 115.67%, and average value of 130.01% with a standard deviation of 40.11% for all the four overlaps considered in the simulation study. The percent relative bias are negligible and lie between -0.210% and 0.193%, -0.191% and 0.176%, -0.175% and 0.161%, and between -0.161% and 0.148% for the small, medium, large and complete overlaps respectively. No doubt the percent RB differs with overlaps but not the percent relative efficiency value, because the value of percent relative bias is small and does not reflect any change in the value of percent relative efficiency up to two decimal places reported in Table 4.1.

It seems that the total of the overlap in two frames is not as important as one could think without doing such a simulation study, which seems a first fruitful observation we have made

Table 4.1: Descriptive statistics of the percent relative efficiency and percent relative BIAS of the shrinkage estimator

			Mean	Std.Dev.	Min	Med	Max		Min	Max
Overlap	**Y_{ab}**	**Freq.**	**Relative Efficiency**						**Relative BIAS**	
Small	500	1965	130.01	40.11	101.01	115.67	430.71		-0.210	0.193
Medium	1000	1965	130.01	40.11	101.01	115.67	430.71		-0.191	0.176
Large	1500	1965	130.01	40.11	101.01	115.67	430.71		-0.175	0.161
Complete	2000	1965	130.01	40.11	101.01	115.67	430.71		-0.161	0.148

Table 4.2: Descriptive statistics of the percent relative efficiency and percent relative BIAS of the shrinkage estimator

		Mean	Std. Dev.	Min	Med	Max		Min	Max
$Corr\left(\hat{Y}_a, \hat{Y}_{ab}\right)$	Freq.	Relative Efficiency						Relative BIAS	
-0.9	700	137.37	41.91	101.24	118.36	341.97		-0.163	0.175
-0.8	700	125.10	24.61	101.05	115.93	234.10		-0.162	0.187
-0.7	544	123.32	24.45	101.04	114.75	233.89		-0.165	0.149
-0.6	616	124.01	35.27	101.08	110.53	320.03		-0.149	0.157
-0.5	916	123.71	30.21	101.01	113.49	319.93		-0.151	0.142
-0.4	908	129.13	37.97	101.03	115.42	349.63		-0.185	0.138
-0.3	1104	131.84	47.36	101.02	115.22	409.47		-0.158	0.149
-0.2	1136	132.84	47.37	101.01	114.84	430.71		-0.210	0.161
-0.1	1236	135.65	45.37	101.05	119.07	401.54		-0.165	0.193

Table 4.3: Descriptive statistics of the percent relative efficiency and percent relative BIAS of the shrinkage estimator

		Mean	Std. Dev.	Min	Med	Max		Min	Max
$Corr\left(\hat{Y}_b, \hat{Y}_{ba}\right)$	Freq.	Relative Efficiency						Relative BIAS	
-0.9	868	170.33	75.34	101.08	141.29	430.71		-0.210	0.160
-0.8	868	145.77	46.94	101.05	126.85	298.09		-0.153	0.172
-0.7	844	130.13	29.56	101.03	119..54	232.37		-0.165	0.193
-0.6	852	121.79	21.93	101.05	113.52	198.02		-0.165	0.165
-0.5	836	121.09	22.54	101.01	112.02	253.44		-0.163	0.144
-0.4	800	118.55	20.99	101.02	111.40	273.54		-0.152	0.154
-0.3	932	119.63	20.64	101.01	112.53	221.51		-0.142	0.162
-0.2	892	122.37	30.18	101.03	113.49	341.97		-0.162	0.159
-0.1	968	121.09	24.73	101.02	112.83	254.12		-0.165	0.187

from our simulation study results. The second important observation we have made from the simulation study is that the most important factor playing a role in the percent relative efficiency is the value of correlation coefficient between the estimators of overlap total Y_{ab} and non-overlap totals Y_a and Y_b, that is, the values of $\text{Corr}\left(\hat{Y}_a, \hat{Y}_{ab}\right)$ and $\text{Corr}\left(\hat{Y}_b, \hat{Y}_{ba}\right)$. Table 4.2 and Table 4.3 give frequency of the percent relative efficiency values being more than 101% along with median, maximum values, average values and standard deviation for different values of $\text{Corr}\left(\hat{Y}_a, \hat{Y}_{ab}\right)$ and $\text{Corr}\left(\hat{Y}_b, \hat{Y}_{ba}\right)$. Note that freq is used to represent the frequency the proposed estimator shows better performance for a particular choice of parameters in the first column.

From Tables 4.2 and 4.3, we found that the values of percent relative efficiency and relative bias are quite different for different values of $\text{Corr}\left(\hat{Y}_a, \hat{Y}_{ab}\right)$ and $\text{Corr}\left(\hat{Y}_b, \hat{Y}_{ba}\right)$. The third observation from the simulation study is that the frequency of the value of percent relative efficiency being more than 101% is different for a particular choice $\text{Corr}\left(\hat{Y}_a, \hat{Y}_{ab}\right)$ than the same particular choice of $\text{Corr}\left(\hat{Y}_b, \hat{Y}_{ba}\right)$. For example, from Table 4.2, if $\text{Corr}\left(\hat{Y}_a, \hat{Y}_{ab}\right) = -0.9$ then there are 700 cases where the RE value remains more than 101% leading to average value of the percent relative efficiency of 137.37% with a standard deviation of 41.91%, median value of 118.36% and maximum value of 341.97%. The value percent relative bias lies between -0.163% and 0.175%. On the other hand, from Table 4.3, if $\text{Corr}\left(\hat{Y}_a, \hat{Y}_{ab}\right) = -0.9$ then there are 868 cases where the RE value remains more than 101% leading to average value of the percent relative efficiency of 170.33% with a standard deviation of 75.34%, median value of 141.29% and maximum value of 430.71%. The value percent relative bias lies between –0.210% and 0.160%.

The Tables 4.4 and 4.5 give the minimum, median and maximum values of the optimum values of α (in (1.16)) and the same descriptive statistics for the optimum values of λ_1, λ_2 and λ_3 (in (3.8)–(3.10)) for different values of $\text{Corr}\left(\hat{Y}_a, \hat{Y}_{ab}\right)$

Table 4.4: Optimum values of unknown parameters computed and used in the simulation study

$Corr\left(\hat{Y}_a, \hat{Y}_{ab}\right)$	Freq.	α		λ_1		λ_2		λ_3	
		Min	**Max**	**Min**	**Max**	**Min**	**Max**	**Min**	**Max**
-0.9	868	-0.192	0.750	0.229	1.979	0.216	1.618	-0.364	0.916
-0.8	868	-0.106	0.801	0.249	2.007	0.194	1.602	-0.249	1.104
-0.7	844	-0.019	0.837	0.276	2.014	0.186	1.580	-0.138	1.292
-0.6	852	0.030	0.903	0.300	2.005	0.192	1.558	-0.004	1.292
-0.5	836	0.108	0.942	0.321	2.021	0.179	1.542	0.055	1.327
-0.4	800	0.151	0.909	0.342	2.009	0.189	1.528	0.110	1.327
-0.3	932	0.163	1.044	0.360	1.956	0.234	1.511	0.164	1.392
-0.2	892	0.222	1.092	0.382	2.026	0.178	1.492	0.204	1.437
-0.1	968	0.250	1.192	0.400	2.035	0.175	1.479	0.243	1.501

Table 4.5: Optimum values of unknown parameters computed and used in the simulation study

$Corr\left(\hat{Y}_b, \hat{Y}_{ba}\right)$	Freq.	α		λ_1		λ_2		λ_3	
		Min	**Max**	**Min**	**Max**	**Min**	**Max**	**Min**	**Max**
-0.9	700	0.309	1.192	0.347	2.035	0.175	1.324	0.396	1.502
-0.8	700	0.207	1.143	0.393	1.922	0.263	1.484	0.152	1.280
-0.7	544	0.201	0.954	0.383	1.829	0.337	1.492	0.142	1.127
-0.6	616	0.193	0.919	0.282	1.743	0.404	1.577	-0.019	0.990
-0.5	916	0.165	0.908	0.256	1.673	0.463	1.592	-0.076	0.901
-0.4	908	0.140	0.879	0.247	1.607	0.512	1.603	-0.149	0.869
-0.3	1104	0.044	0.837	0.241	1.558	0.551	1.607	-0.210	0.811
-0.2	1136	-0.106	0.793	0.233	1.526	0.580	1.614	-0.291	0.782
-0.1	1236	-0.193	0.750	0.228	1.489	0.609	1.618	-0.364	0.742

and $\text{Corr}\left(\hat{Y}_b, \hat{Y}_{ba}\right)$. Similar tables can be developed by an agency or a company who are using dual frame survey technique in repeated surveys, and a good guess of the shrinkage parameters can used in estimation stage for a current survey based on a good guess about the values of correlation coefficients $\text{Corr}\left(\hat{Y}_a, \hat{Y}_{ab}\right)$ and $\text{Corr}\left(\hat{Y}_b, \hat{Y}_{ba}\right)$.

We also presented our findings with dot plots. Although Fig 4.1 is very much self explanatory for different situations, but we also explain it in brief as follows. The four dot plots for small, medium, large and complete overlaps indicate that there is no difference between the individual percent relative efficiency values.

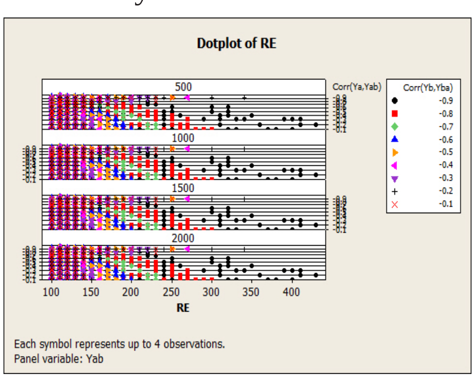

Fig. 4.1. Dot plots showing relationship between the RE values and other parameters

When the value of $\text{Corr}\left(\hat{Y}_b, \hat{Y}_{ba}\right)$ {which is $\text{Corr}\left(\hat{Y}_b, \hat{Y}_{ba}\right)$ in the graph} is close to -0.9 and the value of $\text{Corr}\left(\hat{Y}_a, \hat{Y}_{ab}\right)$ {which is $\text{Corr}\left(\hat{Y}_a, \hat{Y}_{ab}\right)$ in the graph} is close to -0.1, then the proposed shrinkage estimator shows higher value of the percent relative efficiency. In the same way, other values of the percent relative efficiency for different combinations of the correlation coefficients can be interpreted.

Fig. 4.2 has been devoted to visualize a relationship between the percent relative bias and other parameters involved in the estimation process. The relative bias is negligible ranging from -0.20% to 0.20% and hence is not of much interest of discussion left here.

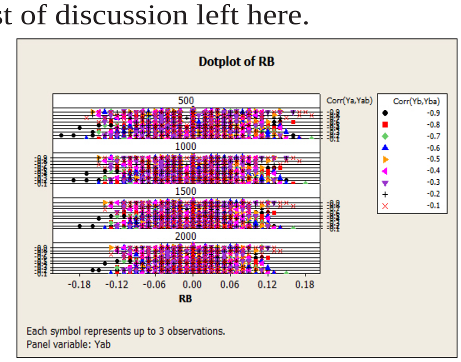

Fig. 4.2. Dot plots showing relationship between the RB values and other parameters

Although relative bias is negligible, but we found a very interesting pattern between the percent relative efficiency values and the percent relative bias values as shown in Fig. 4.3.

From Fig. 4.3, the first observation we made is again that the results are not functions the four types of overlaps, but are functions of values of correlation coefficient values $\text{Corr}\left(\hat{Y}_a, \hat{Y}_{ab}\right)$ and $\text{Corr}\left(\hat{Y}_b, \hat{Y}_{ba}\right)$. Another interesting observation is that when the value of RB is close to zero then the value of the percent relative efficiency remains higher which is quite interesting result.

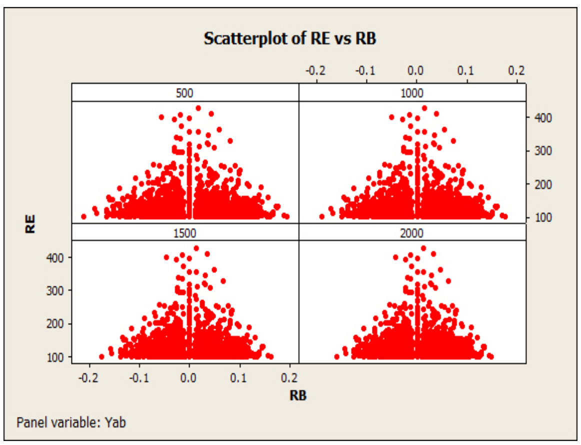

Fig. 4.3. Scatter plots showing relationship between RB and RE values for four levels of overlaps considered

Remarks (a): The proposed shrinkage estimator depends on unknown parameters λ_i, $i = 1, 2, 3$ in the same way as the pioneer Hartley's estimator depends only on unknown parameter α in (1.14). Note that these parameters can be easily replaced by estimates in any shrinkage type estimator as shown in Mangat *et al.* (1991).

(b): One obvious improvement of Fuller and Burmeister (1972) can be seen in a class room exercise:

$$\hat{Y}_{\text{FB(new)}} = \delta_1\hat{Y}_a + \delta_2\hat{Y}_b + \delta_3\left(\hat{N}_{ab} - \hat{N}_{ba}\right) + \delta_4\hat{Y}_{ab} + (1-\delta_4)\hat{Y}_{ba}$$

where $\delta_j, j = 1, 2, 3, 4$ are constants to be determined such that the mean squared error of the estimator is minimum.

REFERENCES

Fuller, W.A. and Burmeister, L.F. (1972). Estimators for samples selected from two overlapping frames. *Proceedings of the Social Statistics Section, American Statistical Association*, 245-249.

Haines, D.E. and Pollock, K.H. (1998). Estimating the number of active and successful bald eagle nests: An application of the dual frame method. *J. Environ. Ecol. Statist.*, **5**, 245-256.

Hartley, H.O. (1962). Multiple frame surveys. *Proceedings of the Social Statistics Section, American Statistical Association*, 203-206.

Hartley, H.O. (1974). Multiple frame methodology and selected applications. *Sankhya*, C, **36**, 99-118.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc*., **47**, 663-685.

Lohr, S. (2011). Alternative survey sample designs: Sampling with multiple overlapping frames. *Survey Methodology*, **37(2)**, 197-213.

Mangat, N.S., Singh, S. and Singh, R. (1991). An alternative estimator in randomized response technique. *Aligarh J. Statist.*, **11**, 75-80.

Searls, D.T. (1964). The utilization of known coefficient of variation in the estimation procedure. *J. Amer. Statist. Assoc.*, **59**, 1125-1126.

APPENDIX I

Fortran codes used in the simulation study

Dual Frame Shrankage Estimator

```
      USE NUMERICAL_LIBRARIES
      IMPLICIT NONE
      INTEGER NS
      REAL  VYa,VYb,VYab,VYba,CYaYab,CY
     1bYba,Ya,Yb,Yab
      REAL VHART, A, B, C, D, E,AL1, AL2,
     1AL3, AMSEP, RE
      REAL ALPH,RB,ARB,YTOT, RHOYaYab,
     1RHOYbYba
      CHARACTER*20 OUT_FILE
      CHARACTER*20 IN_FILE
      WRITE(*,'(A)') 'NAME OF THE OUTPUT
      FILE'
      READ(*,'(A20)') OUT_FILE
      OPEN(42,          FILE=OUT_FILE,
      STATUS='UNKNOWN')
      DO 8888 Yab = 500, 2001, 500
      DO 8888 VYa = 50, 151, 50
      DO 8888 VYb = 50, 151, 50
      DO 8888 VYab = 50, 151, 50
      DO 8888 VYba = 50, 151, 50
      DO 8888 RHOYaYab = -0.9, -0.05, 0.1
      DO 8888 RHOYbYba = -0.9, -0.05, 0.1
      CYaYab = RHOYaYab*SQRT(VYa*VYab)
      CYbYba = RHOYbYba*SQRT(VYb*VYba)
      Ya = 2000
      Yb = 2500
      VHART  =  VYa  +  VYb  +  VYba  +
      12.0*CYbYba
      1 - ( VYba + CYbYba – CYaYab)**2/(VYab
      1+ VYba )
      A  =  (Ya**2  +  VYa)*(VYab  +  VYba)  –
      CYaYab**2
      B = Ya*Yb*( VYab + VYba ) + CYaYab *
      CYbYba
      C = (VYab + VYba) *(Ya**2 + Ya*Yb) –
      1CYaYab*VYba
      D = (VYab + VYba)*(Yb**2 + VYb ) –
      1CYbYba
      E = (VYab + VYba)*(Yb**2 + Ya*Yb –
      CYbYba) + VYba * CYbYba
      AL1 = (C*D-B*E)/(A*D-B**2)
      AL2 = (A*E-B*C)/(A*D-B**2)
      AL3 = ( VYba*(A*D – B**2) – (C*D –
      1B*E)*CYaYab + (A*E-B*C)*CYbYb )/
      1( (A*D-B**2)*(VYab + VYba) )
      AMSEP = (AL1-1)**2*Ya**2 + (AL2-
      1)**2*Yb**2+2.0*(AL1-1)*(AL2-
      1)*Ya*Yb + AL1**2* VYa + AL2**2*VYb
      1 + AL3**2 *VYab +(1-AL3)**2 *VYba +
      1 + 2.0*AL1*AL3*CYaYab
      1 + 2.0*AL2*(1-AL3)*CYbYba
      RE = VHART*100/AMSEP
      ALPH = (VYba + CYbYba – CYaYab)/
      1(VYab + VYba)
      YTOT = Ya + Yb + Yab
      RB = ( (AL1-1.)*Ya + (AL2-1.)*Yb)*
      1 100.0/YTOT
      ARB = ABS(RB)
      IF( (RE.GT.101).AND.(ARB.LT.10) )THEN
      WRITE(*,101)Ya,Yb,Yab,VYa,VYb,VY
      1ab,VYba,CYaYab,CYbYba,RE,RB,
      1-ALPH, AL1, AL2, AL3, RHOYaYab,
      RHOYbYba
      WRITE(42,101)Ya,Yb,Yab,VYa,VYb,V
      1Yab,VYba,CYaYab,CYbYba,RE,RB,
      1ALPH, AL1, AL2, AL3, RHOYaYab,
      1RHOYbYba
101   FORMAT (2X, 15(E14.6, 1X), 1X, 2(F8.2,
      11X))
      ENDIF
8888  CONTINUE
      STOP
      END
```