P.K. BEDI (*) - T.J. RAO (**)

Probability proportional to revised sizes with replacement scheme

1. INTRODUCTION

Consider a finite population of N units, $U_1, U_2, \ldots, U_N$ and let y be the study variate taking values Y_i on U_i, $i = 1, 2, \ldots, N$. One of the problems encountered in practice is the estimation of population total $Y = \sum_{i=1}^{N} Y_i$. When a sample $u = (u_1, u_2, \ldots, u_n)$ of size n is selected by Simple Random Sampling With Replacement (*SRSWR*), an unbiased estimator of Y is given by

$$\hat{Y}_{SRS} = \frac{N}{n} \sum_{i=1}^{n} y_i \tag{1.1}$$

with variance

$$V\left(\hat{Y}_{SRS}\right) = \frac{1}{n}\left[N\sum_{i=1}^{N} Y_i^2 - Y^2\right] \tag{1.2}$$

(*) Formerly of Department of Statistics, University of Rajasthan, Jaipur - 302004, India (Dr. Bedi passed away in November '95).

(**) Stat. Math. Division, Indian Statistical Institute, Calcutta - 700035, India.

In most of the survey situations, some additional information on U_i called auxiliary information is available. Hansen and Hurwitz (1943) were the first to popularize the use of this additional information which takes values X_i on U_i in selecting the units with unequal probabilities. They recommended that units in the sample u be selected with Probability Proportional to Size *i. e.* X_i and With Replacement (*PPSWR*) sampling scheme. An unbiased estimator of Y is given by

$$\hat{Y}_{PPS} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i} \tag{1.3}$$

with variance

$$V\left(\hat{Y}_{PPS}\right) = \frac{1}{n} \left[\sum_{i=1}^{N} \frac{Y_i^2}{p_i} - Y^2 \right] \tag{1.4}$$

where $p_i = X_i / X$ and $X = \sum_{i=1}^{N} X_i$.

Raj (1954) made a direct comparison of $\hat{Y}_{PPS}$ with $\hat{Y}_{SRS}$ and obtained a condition in favour of $\hat{Y}_{PPS}$ as

$$\sum_{i=1}^{N} \left(X_i - \overline{X}\right) Y_i^2 / X_i > 0 \tag{1.5}$$

where $\overline{X}$ is the population mean of the auxiliary variable, which is difficult to verify in practice. Since the exact comparison between the two sampling strategies is not possible, recourse was taken to compare the expected variances under an assumed Super Population Model (*SPM*). In the literature, an often used *SPM*, which also takes into account the intercept, is

$$\left.\begin{array}{c} y_i = \alpha + \beta p_i + e_i \;, \; i = 1, 2, \ldots, N, \\ E\left(e_i \mid p_i\right) = 0 \;,\; E\left(e_i^2 \mid p_i\right) = \sigma^2 p_i^g \\ E\left(e_i e_i \mid p_i, p_i\right) = 0 \;,\; \sigma^2 > 0, \; g \geq 0 \end{array}\right] \tag{1.6}$$

where E (·) denotes the average over all finite populations that can be drawn from the super population. Henceforth this *SPM* will be denoted by model M1 and when $\alpha = 0$ we shall call it as model M2.

Under the above model M1, it is known (Murthy (1967)) that *PPSWR* scheme will be more efficient than *SRSWR* scheme if

$$\sigma^2 \ \text{cov}\left(X_i^{g-1},\ X_i\right) \ > \ \alpha^2 \left(\frac{\bar{X}-\tilde{X}}{\tilde{X}}\right) - \beta^2\sigma_x^2 \tag{1.7}$$

Where $\tilde{X}$ and σ_x^2 are the harmonic mean and variance of the auxiliary variable respectively. This condition is also very difficult to verify in practice. Earlier, Raj (1958) considered the model M2 in which the condition (1.7) is simply

$$\sigma^2 \ \text{cov}\left(X_i^{g-1},\ X_i\right) \ > -\beta^2\sigma_x^2 \tag{1.8}$$

which will be automatically satisfied if $g > 1$ since $\sigma_x^2 > 0$ and cov $(X_i^{g-1}, X_i) > 0$. However, for $0 < g < 1$ the condition (1.8) needs to be satisfied.
Ray (1954) also considered a deterministic model $Y_i = \alpha + \beta X_i$ and then the expression (1.7) reduces in this situation to

$$\frac{\bar{X}-\tilde{X}}{\tilde{X}\ \sigma_x^2} < \frac{\beta^2}{\alpha^2}. \tag{1.9}$$

This condition is likely to be satisfied when regression line is close to the origin. Also the form of $\hat{Y}_{PPS}$ supports this argument as exact proportionality between y and x makes its variance zero. However, in view of the condition (1.9) it can be said that the estimator $\hat{Y}_{SRS}$ will be better than the estimator $\hat{Y}_{PPS}$, if the line of regression is far away from the origin. Thus linearity between y and x is not a sufficient condition for *PPSWR* scheme to be better than *SRSWR* scheme. So in practice it is very difficult to have an idea whether *PPSWR* sampling scheme can be preferred over *SRSWR* scheme, unless the other conditions as mentioned above hold.

In order to overcome this problem, Reddy and Rao (1977) suggested that the sample be selected by a probability proportional to revised sizes scheme and with replacement. The revised sizes are obtained through a location shift in auxiliary variable as $X_i' = X_i + (1 - L)\bar{X} / L$, where $0 < L < 1$. This can also be treated as a compromise selection probability between *PPX* and SRS leading to new

$$p_i' = Lp_i + (1 - L) / N \tag{1.10}$$

which gives the above transformation (Rao (1993)). The unbiased estimator of Y in this case is

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i'} \tag{1.11}$$

with variance

$$V(\hat{Y}) = \frac{1}{n} \left[\sum_{i=1}^{N} \frac{Y_i^2}{p_i'} - Y^2 \right]. \tag{1.12}$$

Reddy and Rao (1977) proved that the estimator $\hat{Y}$ will be better than the worse of the estimators $\hat{Y}_{SRS}$ and $\hat{Y}_{PPS}$ in $0 < L < 1$. They further recommended the value of $L = k$, which is the value of the optimum transformation parameter for ratio estimator with *SRSWOR* scheme (c.f. Reddy (1974)), where $k = \beta/R$; β is the regression coefficient of y on x and $R = Y/X$. Therefore, the estimator $\hat{Y}$ and $V(\hat{Y})$ reduce to

$$\hat{Y}' = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{kp_i + (1-k)/N} \tag{1.13}$$

and

$$V(\hat{Y}') = \frac{1}{n} \left[\sum_{i=1}^{N} \frac{Y_i^2}{kp_i + (1-k)/N} - Y^2 \right] \tag{1.14}$$

respectively. Singh *et al.* (1983) suggested the construction of a condensed variable Q as a linear function of several auxiliary variables obtained by simultaneously maximizing the correlation coefficient between y and Q and minimizing the intercept of line of regression y on Q with the restriction that $V(Q)$ is unity. Further, under the modified super population model obtained from model M2 by replacing p by Q they compare the efficiency of the estimator based on the condensed variable in the *PPSWR* scheme with that of the corresponding conventional estimator based on auxiliary variable having maximum correlation with y. In case of one auxiliary variable, the condensed variable Q reduces to the transformed auxiliary variable X_i' with $L = k$ and thus we have the estimator $\hat{Y}'$ through a different interpretation.
Reddy and Rao (1977) and Rao (1993) studied the modified *PPSWR* scheme with $0 < L < 1$ whereas its application in practice requires that $p_i' > 0$ which gives a wider range for L as

$$-1/(Np_{\max}-1) < L < 1/(1-Np_{\min}) \tag{1.15}$$

where $p_{\min}$ and $p_{\max}$ are the minimum and maximum initial probabilities of selection.

Therefore, an attempt has been made, without any restriction on the value of L or k, to obtain some results regarding the comparison of the estimators $\hat{Y}$ or $\hat{Y}'$ with that of estimators $\hat{Y}_{SRS}$ and $\hat{Y}_{PPS}$ in section 2 on the basis of expected variances of the estimators obtained firstly under the super population model M2 and secondly under the modified super population model M3 considered by Singh *et al.* (1983). Near – optimum value of L which minimizes $V(\hat{Y})$ is derived in section 3. The last section gives an empirical illustration to observe the gain in efficiency by using the near – optimum value of L or k and for deviations from it.

2. Model based comparison of estimators

In this section, we first give two lemmas which will be useful for comparison of the estimators $\hat{Y}$, $\hat{Y}_{SRS}$ and $\hat{Y}_{PPS}$ under models M2 and model M3 considered by Singh *et al.* (1983).

Lemma 2.1 (Royall (1970)): Let $0 \le b_1 \le b_2 \le, \ldots, \le b_m$ and $c_1 \le c_2 \le, \ldots, \le c_m$ satisfying $\sum_{i=1}^{m} c_i \ge 0$. Then $\sum_{i=1}^{m} b_i\, c_i \ge 0$.

Lemma 2.2: Let $b_1 \ge b_2 \ge, \ldots, \ge b_m \ge 0$ and $c_1 \ge c_2, \ldots, \ge c_m$ satisfying $\sum_{i=1}^{m} c_i \ge 0$. Then $\sum_{i=1}^{m} b_i\, c_i \ge 0$.

The expected variances of the estimators $\hat{Y}_{SRS}$, $\hat{Y}_{PPS}$ and $\hat{Y}$ under the super population model M2 are

$$n E\, V\left(\hat{Y}_{SRS}\right) = \beta^2 \left[N \sum_{i=1}^{N} p_i^2 - 1 \right] + \sigma^2 (N-1) \sum_{i=1}^{N} p_i^g$$

$$n E\, V\left(\hat{Y}_{PPS}\right) = \sigma^2 \sum_{i=1}^{N} p_i^{g-1} \left(1 - p_i\right)$$

and

$$n E\, V\left(\hat{Y}\right) = \beta^2 \left[\sum_{i=1}^{N} \frac{p_i^2}{p_i'} - 1 \right] + \sigma^2 \sum_{i=1}^{N} p_i^g \left(\frac{1}{p_i'} - 1 \right)$$

respectively. Now, we compare the estimator $\hat{Y}$ with $\hat{Y}_{SRS}$ under the super population model M2 in the following theorem:

THEOREM 2.1: Under the super population model M2 the sufficient condition that the estimator $\hat{Y}$ has smaller expected variance than the estimator $\hat{Y}_{SRS}$ for $0 < L < 2 / (2 - Np_{\min})$ is $g > (Lp_{\max}) / p'_{\max}$.

Proof: The difference between the expected variance of the estimators $\hat{Y}_{SRS}$ and $\hat{Y}$ can be written as

$$n\left[E\, V\left(\hat{Y}_{SRS}\right) - E\, V\left(\hat{Y}\right)\right] = \beta^2 \sum_{i=1}^{N} b_i c_i + \sigma^2 \sum_{i=1}^{N} b_i' c_i \qquad (2.1)$$

where $c_i = Np'_i - 1$, $b_i = p_i^2 / p'_i$ and $b'_i = p_i^g / p'_i$. Now, because $\sum c_i = 0$ and c_i is an increasing function of p_i for $L > 0$ and so is b_i as long as $0 < L < 2 / (2 - Np_{\min})$. The sufficient condition that b'_i should also be an increasing function of p_i is

$$g > Lp_i / p'_i$$

Thus, in view of Roayall's lemma 2.1 both parts of the expression (2.1) are positive with highest of the upper limit of g being for suffix i = max for some i and hence the theorem.

Remark 2.1: For the choice of $-2 / (Np_{max} - 2) < L < 0$ the p'_i remains positive and the same is true with the first term of (2.1) by using lemma 2.2 but the condition of positivity of second term comes out to be $g < Lp_i / p'_i$ which rarely happens in practice as $0 \le g \le 2$. Thus the comparison between $\hat{Y}$ and $\hat{Y}_{SRS}$ is inconclusive for the choice of $L < 0$.

Now, we will compare the estimator $\hat{Y}$ with $\hat{Y}_{PPS}$ under the super population model M2 in the following theorem:

THEOREM 2.2: Under the super population model M2 the estimator $\hat{Y}_{PPS}$ has smaller expected variance than the estimator $\hat{Y}$ for $g = 2$. When $g \neq 2$, the superiority of $\hat{Y}_{PPS}$ over $\hat{Y}$ still holds for values of $g > 1 + (Lp_{max})/p'_{max}$ when $L < 1$ but for $L > 1$ it holds for $g < 1 + (Lp_{min})/p'_{min}$.

Proof: The difference between the expected variances o the estimator $\hat{Y}_{PPS}$ and $\hat{Y}$ can be written as

$$n\left[E\ V\left(\hat{Y}\right) - E\ V\left(\hat{Y}_{PPS}\right)\right] = \beta^2 \sum_{i=1}^{N} \left(\frac{p_i}{\sqrt{p'_i}} - \sqrt{p'_i} \right)^2 + \sigma^2 \sum_{i=1}^{N} d_i c'_i \qquad (2.2)$$

where $c'_i = p_i - p'_i$ and $d_i = p_i^{g-1} / p'_i$. Now because $\Sigma c'_i = 0$ and c'_i is an increasing function of p_i as long as $L < 1$. One can show $\Sigma d_i c'_i > 0$ by using lemma 2.1 provided d_i is also an increasing function of p_i. A sufficient condition for this is that first derivative of d_i with respect to p_i is greater than zero which gives $g > 1 + Lp_i / p'_i$ with the highest of upper limit for suffix i = max for some i.

We further observe that c'_i is decreasing function of p_i for $L > 1$. In view of lemma 2.2 we can have $\Sigma d_i c'_i > 0$ provided d_i is also decreasing function of p_i. A sufficient condition for this is $g < 1 + Lp_i / p'_i$ with lowest of the upper limit for suffix i = min for some i.

At $g = 2$ the expression (2.2) reduces to

$$n\left[E\ V\left(\hat{Y}\right) - E\ V\left(\hat{Y}_{PPS}\right)\right] = \left(\beta^2 + \sigma^2\right) \sum_{i=1}^{N} \left(\frac{p_i}{\sqrt{p'_i}} - \sqrt{p'_i} \right)^2$$

which is always positive and hence the theorem.

Remark 2.2: The results of theorems 2.1 and 2.2 are also applicable to the estimator $\hat{Y}'$ if we replace L by k.

Now, the super population model M3, considered by Singh *et al.* (1983) is reproduced below for ready reference in the case of one auxiliary variable, viz.,

$$\left.\begin{array}{l} Y_i = \beta Q_i + e_i \ , \ \ i = 1,2,\ldots,N \ , \\ E(e_i | Q_i) = 0 \ , \ E(e_i^2 | \ Q_i) = \sigma^2 Q_i^g \\ E(e_i e_j | Q_i, Q_j) = 0 \ , \ \sigma^2 > 0, \ \ g \geq 0 \end{array}\right] \tag{2.3}$$

where $Q_i = kp_i + (1 - k) / N$ with $\Sigma Q_i = 1$. Under this model the expected variances of the estimators $\hat{Y}_{SRS}$, $\hat{Y}_{PPS}$ and $\hat{Y}'$ are

$$nE\ V(\hat{Y}_{SRS}) = \beta^2 \left[N \sum_{i=1}^{N} Q_i^2 - 1 \right] + \sigma^2 (N-1) \sum_{i=1}^{N} Q_i^g$$

$$nE\ V(\hat{Y}_{PPS}) = \beta^2 \left[\sum_{i=1}^{N} \frac{Q_i^2}{p_i} - 1 \right] + \sigma^2 \sum_{i=1}^{N} Q_i^g \left(\frac{1}{p_i} - 1 \right)$$

and

$$nE\ V(\hat{Y}') = \sigma^2 \sum_{i=1}^{N} Q_i^{g-1} (1 - Q_i)$$

respectively. Next we compare the estimaor $\hat{Y}'$ with $\hat{Y}_{PPS}$ in the following theorem:

THEOREM 2.3: Under the model specified by (2.3) the estimator $\hat{Y}'$ has always smaller expected variance then the estimator $\hat{Y}_{PPS}$ at $g = 2$. When $g \neq 2$ the superiority of $\hat{Y}'$ over $\hat{Y}_{PPS}$ still holds for values of $g > 1 + Q_{max} / p_{max}$ when $k > 1$ but for $k < 1$ it holds for $g < 1 + Q_{min}/p_{min}$.

Proof: The difference between the expected variances of the estimator $\hat{Y}'$ and $\hat{Y}_{PPS}$ can be written as

$$n\left[E\ V\left(\hat{Y}_{PPS}\right)-E\ V\left(\hat{Y}'\right)\right]=\beta^2\sum_{i=1}^{N}\left(\frac{Q_i}{\sqrt{p_i}}-\sqrt{p_i}\right)^2 + \sigma^2\sum_{i=1}^{N} d_i'' c_i'' \qquad (2.4)$$

where $d_i'' = Q_i^{g-1} / p_i$ and $c_i'' = Q_i - p_i$. Now, because $\sum c_i'' = 0$ and c_i'' is an increasing (decreasing) function of p_i for k greater (less) than 1, to show $\sum d_i''\ c_i'' > 0$ in view of lemma 2.1 a sufficient condition is that d_i'' is also an increasing function of p_i giving $g > 1 + Q_i / p_i$ with highest of the upper limit at suffix i = max for some i.
To have $\sum d_i''\ c_i'' > 0$ in view of lemma 2.2 when $k < 1$, a sufficient condition that d_i'' is also decreasing function of p_i yields $g < 1 + Q_i / p_i$ which is lowest when suffix i = min for some i.

When $g = 2$ the expression (2.4) reduces to

$$n\left[E\ V\left(\hat{Y}_{PPS}\right)-E\ V(Y')\right]=\left(\beta^2+\sigma^2\right)\sum_{i=1}^{N}\left(\frac{Q_i}{\sqrt{p_i}}-\sqrt{p_i}\right)^2$$

which is always positive and hence the theorem.

Now, we will compare the estimator $\hat{Y}'$ with estimator $\hat{Y}_{SRS}$ in the following theorem:

THEOREM 2.4: Under the model M3 specified by (2.3) the estimator $\hat{Y}'$ has smaller expected variance than the estimator $\hat{Y}_{SRS}$ for $g > 1$ and for $g < 1$ if

$$\sigma^2 \text{ cov } (Q_i^{g-1}, p_i) > -\ \beta^2\ \sigma_Q^2.$$

Proof: The difference between the expected variance of the estimator $\hat{Y}_{SRS}$ and $\hat{Y}'$ is

$$n\left[E\ V\left(\hat{Y}_{SRS}\right)-E\ V\left(\hat{Y}'\right)\right]=\beta^2\left(N\sum_{i=1}^{N}Q_i^2-1\right)+\sigma^2\sum_{i=1}^{N}Q_i^{g-1}\left(Np_i-1\right)$$

which can be rewritten as

$$n\left[E\,V\left(\hat{Y}_{SRS}\right)-E\,V\left(\hat{Y}'\right)\right]=N^2\sigma_Q^2\beta^2+\sigma^2N^2\operatorname{cov}\left(Q_i^{g-1},p_i\right) \quad (2.5)$$

where σ_Q^2 is the variance of transformed variable Q. For $g > 1$ Q_i^{g-1} is an increasing function of p_i and so cov $(Q_i^{g-1}, p_i) > 0$ and thus the right hand side of (2.5) will be positive, whereas for $g < 1$, it is so if

$$\sigma^2 \text{ conv } (Q_i^{g-1}, p_i) > -\beta^2\sigma_Q^2$$

and hence the theorem.

3. Near-optimum value of transformation parameter

In this section, we obtain a near-optimum value of L which minimizes the variance of the estimator $\hat{Y}$ in the following theorem:

THEOREM 3.1: Under the assumption that $|L\,(Np_i - 1)| < 1$, the value of L upto the first order of approximation which minimizes the variance of the estimator $\hat{Y}$, is

$$L_1=\frac{\sum_{i=1}^{N}Y_i^2(Np_i-1)}{2\sum_{i=1}^{N}Y_i^2(Np_i-1)^2} \quad (3.1)$$

and the resulting optimum variance is

$$V_0\left(\hat{Y}\right)=V\left(\hat{Y}_{SRS}\right)-\frac{N\left[\sum_{i=1}^{N}Y_i^2(Np_i-1)\right]^2}{4n\sum_{i=1}^{N}Y_i^2(Np_i-1)^2}. \quad (3.2)$$

Proof: The variance of the estimator $\hat{Y}$ in expression (1.12) can be written as

$$V\left(\hat{Y}\right)=\frac{1}{n}\left[N\sum_{i=1}^{N}Y_i^2\{1+L(Np_i-1)\}^{-1}-Y^2\right]. \quad (3.3)$$

Making the assumption $|L(Np_i - 1)| < 1$ and retaining the terms up to the order of L^2, (3.3) reduces to

$$V_1(\hat{Y}) = \frac{1}{n}\left[N\sum_{i=1}^{N} Y_i^2\left\{1 - L(Np_i - 1) + L^2(Np_i - 1)^2\right\} - Y^2\right]. \quad (3.4)$$

The optimum value of L say L_1 which minimizes $V_1(\hat{Y})$ can be obtained by differentiating $V_1(\hat{Y})$ with respect to L and equating to zero. We have

$$L_1 = \frac{\sum_{i=1}^{N} Y_i^2(Np_i - 1)}{2\sum_{i=1}^{N} Y_i^2(Np_i - 1)^2} = \frac{N^2\operatorname{cov}(Y_i^2, p_i)}{2\sum_{i=1}^{N} Y_i^2(Np_i - 1)^2}$$

It is also noted that the second derivative of $V_1(\hat{Y})$ is always positive. So the minimum value of $V_1(\hat{Y})$ is

$$V_1(\hat{Y})_0 = V(\hat{Y}_{SRS}) - \frac{N\left[\sum_{i=1}^{N} Y_i^2(Np_i - 1)\right]^2}{4n\sum_{i=1}^{N} Y_i^2(Np_i - 1)^2}$$

and hence the theorem.

Remark 3.1: An assumption in the theorem 3.1 requires that $|L| <$ Min. $\{(1 - Np_{\min})^{-1}, (Np_{\max} - 1)^{-1}\}$ and so together with (1.15) we can easily have an idea of the upper limit of L. However, for the sign of L we require a prior knowledge about Cov (Y_i^2, p_i).

Remark 3.2: We can rewrite the expression (3.4) of $V_1(\hat{Y})$ as

$$V_1(\hat{Y}) = V(\hat{Y}_{SRS}) - \frac{N}{n}L(2L_1 - L)\sum_{i=1}^{N} Y_i^2(Np_i - 1)^2. \quad (3.5)$$

So the estimator $\hat{Y}$ will be better than the estimator $\hat{Y}_{SRS}$ as long as $0 < |L| < 2|L_1|$.

Remark 3.3: The difference between $V(Y_{PPS})$ and $V_1(Y)$ is

$$V\left(\hat{Y}_{PPS}\right)-V_1\left(\hat{Y}\right)=\left[V\left(\hat{Y}_{PPS}\right)-V\left(\hat{Y}_{SRS}\right)\right]+\frac{N}{n}L\left(2L_1-L\right)\sum_{i=1}^{N}Y_i^2\left(Np_i-1\right)^2.$$

For $0 < |L| < 2\,|L_1|$ the estimator $\hat{Y}$ will always be superior than estimator $\hat{Y}_{PPS}$ if the estimator $\hat{Y}_{SRS}$ is more efficient than the estimator $\hat{Y}_{PPS}$. When the estimator $\hat{Y}_{PPS}$ is more efficient than the estimator $\hat{Y}_{SRS}$ then still there is a scope that $\hat{Y}$ will be better than $\hat{Y}_{PPS}$ but it can not be verified in practice easily.

4. Empirical illustration

To study the behaviour of the estimators $\hat{Y}_{SRS}$, $\hat{Y}_{PPS}$ and $\hat{Y}$ empirically, we consider the well known populations of Yates and Grundy (1953), Cochran (1977) and Amahia *et al.* (1989). For the sake of ready reference these populations are given in the table 4.1. The percentage efficiency of the estimator $\hat{Y}$ which respect to $\hat{Y}_{SRS}$ for the true value of k equal to β/R and L_1, the optimum choice of L given by (3.1) as well as deviations from it is given in the table 4.2. As we know that exact proportionality between y and p makes the estimator $\hat{Y}_{PPS}$ the best; in table 4.3, the percentage efficiencies of the estimator $\hat{Y}$ with respect to the estimator $\hat{Y}_{PPS}$ are given for only those populations in which y is not nearly proportional to p.

It is clear from the table 4.2 that estimator $\hat{Y}$ is better than the estimator $\hat{Y}_{SRS}$ when $0 < L < 2L_1$ (as exhibited in the tables) and also robust for small departure from the true value of L_1 whose corresponding entries in the tables are denoted by ". Table 4.3 shows that the estimator $\hat{Y}$ will be better than the estimator $\hat{Y}_{PPS}$ as well, if L is near k in which case the estimator $\hat{Y}$ has also maximum gain over the estimator $\hat{Y}_{SRS}$. Therefore a good guess of L_1 in practice can be taken to be the value of k.

It is also observed from the Table 4.3 that when the intercept of line of regression of y on x is negative as the case for populations A and E we require $L > 1$ whereas for populations B, F, G and H the intercept is positive and therefore $L < 1$ to achieve gain over *PPSWR* scheme.

TABLE 4.1

POPULATIONS

Unit number	A x	A y	B y	C y	D x	D y	E y	F y	G x	G y	H x	H y
1	0.1	0.5	0.8	0.2	0.1	0.3	0.3	0.7	25	11	41	36
2	0.2	1.2	1.4	0.6	0.1	0.5	0.3	0.6	32	7	34	47
3	0.3	2.1	1.8	0.9	0.2	0.8	0.8	0.4	14	5	54	41
4	0.4	3.2	2	0.8	0.3	0.9	1.5	0.9	70	27	39	47
5					0.3	1.5	1.5	0.6	24	30	49	49
6									20	6	45	45
7									32	13	41	32
8									44	9	33	37
9									50	14	37	40
10									44	18	41	41
11											47	37
12											39	48

TABLE 4.2

PERCENTAGE EFFICIENCY OF $\hat{Y}$ OVER $\hat{Y}_{SRS}$

L	A	B	C	D	E	F	G	H
0	100	100	100	100	100	100	100	100
0.04	106.71	111.69	106.49	106.74	106.20	104.42	102.90	100.2"
0.045	107.59	113.28	107.34	107.62	107.01	104.91	103.24	100.2'
0.1	117.89	133.14	117.20	117.94	116.47	109.07	106.77	
0.156	129.74	158.84	128.38	129.75	127.28	110.7'	109.81	
0.158	130.19	159.88	128.80	130.20	127.69	110.7"	109.91	
0.2	140.19	184.31	138.06	140.09	136.74	109.92	111.79	
0.29	165.46	258.43	160.76	164.65	159.35	103.13	114.7"	
0.3	168.65	269.20	163.55	167.70	162.18	102.01	114.88	
0.4	205.98	423.50	194.82	202.40	194.88		115.95	
0.5	328.80	734	233.14	246.09	238.15		115.07	
0.6	328.80	1370'	279.41	300.47	297.52		112.41	
0.639	365.58	1700	299.59	324.70	326.94		110.9'	
0.667	396.06	1904"	314.73	343.08	350.91		109.76	
0.7	437.36	2022.8	333.13	365.62	382.80		108.22	
0.77	514.13	1680.6	373.2"	415.36	467.42		104.55	
0.8	614.39	1402.1	390.08	436.6"	512.95		102.82	
0.82	663.4"	1220.7	401.01	450.44	547.29		101.62	
0.84	718.80	1054.7	411.5'	463.73	585.38		100.39	
0.9	937.61	673.54	438.89	498.43	728.65			
0.923	1050.4	569.50	447.01	508.71	798.7"			
1	1636	336	460	525'	1131.2			
1.1	3560.3	182.39	434.53	494.84	2032.9			
1.2	8604.3	104.58	362.27	414.81	4747.5			
1.2857	5777'		279.89	329.93	13007			
1.3636	2025.7		204.03	256.10	15426'			
1.4	1266.1		170.98	225	9615.6			
1.5	375.6			153.12	2436.5			
1.6				100	861.90			

' and " denote the percentage efficiency for the true population value of k (= β/R) and L_1 the optimum choice of L respectively.

TABLE 4.3

PERCENTAGE EFFICIENCY OF $\hat{Y}$ OVER $\hat{Y}_{PPS}$

			Population			
L	A	B	E	F	G	H
0				448.48	111.62	214.88
0.04				468.33	114.85	215.3"
0.045				470.52	115.24	215.3'
0.1				489.17	119.17	214.33
0.156				496.4'	122.57	211.69
0.158				496.4"	122.68	211.56
0.2				493.00	124.79	208.54
0.29				462.55	128"	199.55
0.3				457.53	128.23	198.38
0.4		126.04		395.99	129.43	185.13
0.5		218.45		326.47	128.44	170.16
0.6		407.8'		261.62	125.47	154.64
0.639		505.9		238.92	123.8'	148.63
0.667		566.57"		223.64	122.51	144.37
0.7		602.05		206.73	120.80	139.42
0.77		500.20		174.69	116.70	129.27
0.8		417.30		162.47	114.77	125.08
0.82		363.32		154.79	113.44	122.34
0.84		313.92		147.46	122.06	119.65
0.9		200.46		127.49	107.73	111.90
0.923		169.50		120.57	106.01	109.05
1	100.00	100.00	100.00	100.00	100.00	100.00
1.1	217.62		179.71			
1.2	525.94		419.67			
1.2857	353.1'		1149.8			
1.3636	123.82		1363.6'			
1.4			850.00			
1.5			215.38			

' and " denote the percentage efficiency for the true population value of k (= β/R) and L_1 the optimum choice of L respectively.

REFERENCES

AMAHIA, G.N., CHAUBEY, Y.P. and RAO, T.J. (1989) Efficiency of a new estimator in PPS sampling for multiple characteristics, *Jour. Stat. Plan. Inf.*, 21, pp. 75-84.

COCHRAN, W.G. (1977) *Sampling techniques* (Third edition), John Wiley and Sons, New York.

HANSEN, M.H. and HURWITZ, W.N. (1943) On the theory of sampling from finite populations, *Ann. Math. Stat.*, 14, pp. 333-362.

MURTHY, M.N. (1967) *Sampling theory and methods*, Statistical Publishing Society, Indian Statistical Institute, Calcutta, India.

RAJ, D. (1954) On sampling with probabilities proportional to size, *Ganita*, 5, pp. 175-182.

RAJ, D. (1958) On the relative accuracy of some sampling techniques, *Jour. Amer. Stat. Assoc.*, 53, pp. 98-101.

RAO, T.J. (1993) On certain problems of sampling designs and estimation for multiple characteristics, *Sankhyā*, Series B, 55, pp. 372-384.

REDDY, V.N. (1974) On a transformed ratio method of estimation, *Sankhyā*, Series C, 36, pp. 59-70.

REDDY, V.N. and RAO, T.J. (1977) Modified PPS method of estimation, *Sankhyā*, Series C, 39, pp. 185-197.

ROYALL, R.M. (1970) On finite population sampling theory under certain linear regression models, *Biometrika*, 57, pp. 377-387.

SING, M., KUMAR, P. and CHANDAK, R. (1983) Use of multi-auxiliary variables as a condensed auxiliary variable in selecting a sample, *Comm. Stat.* (TM), 12, pp. 1685-1697.

YATES, F. and GRUNDY, P.M. (1953) Selection without replacement from within strata with probability proportional to size, *Jour. Roy. Stat. Soc.*, Series B, 15, pp. 253-261.

Probability proportional to revised sizes with replacement scheme

SUMMARY

An unbiased estimator based on a modified Probability Proportional to Size With Replacement (PPSWR) scheme is suggested by making use of a linear transformation on the auxiliary variable. The proposed estimator is compared with conventional SRS and PPS with replacement estimators under certain super population models. A near– optimum value of the transformation parameter which minimizes the variance of the estimator is also derived. An empirical illustration is also given to find the efficiency of the proposed estimator over the above cited estimators.

Probabilità proporzionali ad ampiezze rivedute nel campionamento con ripetizione

RIASSUNTO

In questo lavoro si propone uno stimatore corretto basato su di una modifica dello schema di campionamento con probabilità proporzionali alla dimensione con ripetizione (PPSWR) e costruito facendo uso di una trasformazione lineare applicata alla variabile ausiliaria. Lo stimatore proposto viene confrontato con gli usuali stimatori per il campionamento semplice con ripetizione (SRS) e per lo schema con probabilità proporzionali all'ampiezza con ripetizione (PPS) in alcuni modelli di superpopolazioni. Inoltre si ottiene un valore quasi-ottimo del parametro da cui dipende la trasformazione utilizzata e che rende minima la varianza dello stimatore. L'efficienza dello stimatore nella sua classe viene valutata empiricamente.

KEY WORDS

Probability proportional to modified size with replacement scheme; transformed auxiliary variable and superpopulation model.