Thai Journal of Mathematics : (2018) 417-429
Special Issue (ACFPTO2018) on : Advances in fixed point theory towards real world optimization problems

http://thaijmath.in.cmu.ac.th
Online ISSN 1686-0209

A New Ratio Estimator for Population total in the Presence of Nonresponse under Unequal Probability Sampling without Replacement

Chugiat Ponkaew and Nuanpan Lawson[1]

Department of Applied Statistics, Faculty of Applied Science
King Mongkuts University of Technology North Bangkok, Bangkok, Thailand
e-mail : nuanpan.n@sci.kmutnb.ac.th

Abstract : In this paper, a modified ratio estimator for estimating population total in the presence of nonresponse under unequal probability sampling without replacement are proposed under the conditions of with or without response probability. The modified estimator is investigated under a reverse framework with a uniform nonresponse mechanism where the overall sampling fraction is negligible. Theoretical studies show that the proposed estimator is asymptotically unbiased and more efficient than the existing one. We compared the efficiency of the proposed estimator with other estimator through a simulation study. The results showed that for all levels of response probability the proposed estimator had a smaller relative bias and relative root mean square error than other estimator.

Keywords : Uniform nonresponse, Reverse framework, Taylors linearization approach.
2010Mathematics Subject Classification : 62D05.

1 Introduction

The problem of constructing efficient estimators for the population total by using auxiliary variables has been widely discussed by various authors. Cochran [1] proposed a ratio estimator under simple random sampling without replacement when the population total of an auxiliary variable is known. Sisodia and Dwivedi [2], Singh and Kakran [3], Upadhyaya and Singh [4] also suggested ratio-type estimators for estimating the population total when the population coefficient of variation and the kurtosis of the auxiliary variable are known under simple random sampling without replacement. Later, Bacanli and Kadilar [5] extended the estimators of Sisodia and Dwivedi [2], Singh and Kakran [3], and Upadhyaya and Singh [4] to unequal probability sampling without replacement. They also showed that their proposed estimators are more efficient than the classical ratio estimator by using mean square error calculations to compare them.

The ratio estimator has a form of nonlinear function, therefore properties such as mean, variance and mean square error can be obtained by using the Taylor linearization approach. This method was first discussed by Tepping [6] who applied it to an investigation into the variance of nonlinear function. Later, the Taylor linearization approach has been widely discussed by various authors such as Woodruff [7], Wolter [8], Demnati and Rao [9]. Under the Taylor linearization approach, a linear approximation of the nonlinear function is obtained by using Taylor series approximation. Therefore a variance of nonlinear function can be calculated by deriving the variance of the linear approximation.

In the presence of nonresponse the traditional estimators can not be used to estimate population total because a selected sample set is divided between a response and nonresponse set. Therefore Bethlehem [10] proposed a modified Horvitz and Thompson [11]'s estimator which is in a more general form for a random response approach. Srndal and Lundstrm [12] proposed an estimator for estimating the population total under a two-phase framework with an assumption that the response probability is known under unequal probability sampling without replacement. This estimator is like Horvitz and Thompson [11] ' s estimator but uses sample units of study variables in response sets instead of sample sets. Later, Lawson [13] proposed an estimator for estimating population totals under a reverse framework when the overall sampling fraction is negligible under a uniform nonresponse mechanism.

This paper aims to estimate a population total by using a ratio estimator to improve the estimation efficiency under unequal probability sampling without replacement when nonresponse occurs in the study variable only. We proposed to modify the Bacanli and Kadilar [5] 's estimator by using the estimator of Srndal and Lundstrm [12]under a reverse framework with an overall negligible sampling fraction and uniform nonresponse. We discuss notation and framework in section 2. In section 3, we introduce Srndal and Lundstrm [12]'s estimator and associated variance under reverse framework. In section 4, the proposed new ratio estimator and associated variance are discussed. In section 5, we compare the efficiency of the proposed estimator with Srndal and Lundstrm [12]'s estimator. Finally some conclusions are given in section 6.

2 Notation and framework

Consider a finite population $U = \{\ 1, 2, ..., N\}$. Let y_i be the value of random variable y for the ith population unit. Assume that a sample s of size n was selected under unequal probability sampling without replacement. Our aim is to estimate the population total of y defined by $Y = \sum_{i \in U} y_i$. Let x be an auxiliary variable that is highly correlated with the study variable. We assume the population total $X = \sum_{i \in U} x_i$ is known. Let π_i and π_{ij} denote the first and second order of inclusion probabilities defined by $\pi_i = P(i \in s)$ and $\pi_{ij} = P(i \wedge j \in s)$ respectively. Let us define a random variable I_i where $I_i = 1$ if $i \in s$ otherwise $I_i = 0$.

Under nonresponse, let r_i denote the response indicator variable of y_i and be defined by $r_i = 1$ if y_i is observed otherwise $r_i = 0$. Let $\boldsymbol{R} = (r_1, r_2, ..., r_N)'$ denoted be the vector of response indicators. Let p_i denote the response probability as defined by $p_i = P(r_i = 1)$. We also assume that (A) Nonresponse mechanism has uniform nonresponse and (B) that the overall sampling fraction is negligible.

3 Srndal and Lundstrm [12]'s estimator and associated variance under reverse framewrok

In the full response, Horvitz and Thompson [11] proposed an unbiased estimator for estimating population total based on sample element. In the presence of nonresponse, the Horvitz and Thompson [11]'s estimator does not work because some units of the study variable cannot be observed. Therefore Srndal and Lundstrm [12] proposed an estimator based on Horvitz and Thompson [11]'s estimator under a two-phase framework. In this section, we extended Srndal and Lundstrm's [12] estimator to include a reverse framework and defined in Definition 3.1.

Definition 3.1. Assume that (A) holds. Under a reverse framework with unequal probability sampling without replacement where p is known. The Srndal and Lundstrm [12]'s estimator for Y is defined by,

$$\hat{Y}_r^{(1)} = \sum_{i \in s} \frac{r_i y_i}{\pi_i p}. \tag{3.1}$$

Lemma 3.2. *Assume that (A) holds. Under a reverse framework with unequal probability sampling without replacement where p is unknown. The Srndal and Lundstrm [12] 's estimator for Y is defined by,*

$$\hat{Y}_r^{(2)} = \sum_{i \in s} \frac{r_i y_i}{\pi_i \hat{p}}, \tag{3.2}$$

where $\hat{p} = \left(\sum_{i \in s} \frac{r_i}{\pi_i} \right) \left(\sum_{i \in s} \frac{1}{\pi_i} \right)^{-1}$.

Theorem 3.3. *Assume that (A) holds. Under a reverse framework with unequal probability sampling without replacement, the Srndal and Lundstrm [12]'s estimator is asymptotically unbiased estimators of Y .*

Theorem 3.4. *Assume that (A) and (B) remain correct. Under a reverse framework with unequal probability sampling without replacement.*

(i) *If p is known then the variance of $\hat{Y}_r^{(1)}$ is defined by,*

$$V(\hat{Y}_r^{(1)}) \approx \sum_{i\in U} D_{2i} y_i^2 + \sum_{i\in U}\sum_{j\neq i\in U} D_{ij} y_i y_j, \tag{3.3}$$

where $D_{2i} = \frac{1-\pi_i}{\pi_i p}$ and $D_{ij} = \frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}$.

(ii) *If p is unknown then the variance of $\hat{Y}_r^{(2)}$ is given by,*

$$V(\hat{Y}_r^{(2)}) \approx \sum_{i\in U} D_{2i}(y_i - \bar{Y})^2 + \sum_{i\in U} D_i(2y_i - \bar{Y})\bar{Y} + \sum_{i\in U}\sum_{j\neq i\in U} D_{ij} y_i y_j, \tag{3.4}$$

where D_{2i}, D_{ij} are defined in (3.3) and $D_i = \frac{1-\pi_i}{\pi_i}$.

4 The proposed estimator and associated variance

4.1 The proposed estimator

In this section, we proposed a new ratio estimator for estimating the population total when nonresponse occurs only in the study variable. Cochran [1] first proposed a ratio estimator under simple random sampling without replacement when the population totals of an auxiliary variable is known to increase the efficiency of the population total estimator. Later Bacanli and Kadilar [5] also discussed a ratio estimator based on Horvitz and Thompson [11]'s estimator under unequal probability sampling without replacement and defined by,

$$\hat{Y}_R = \frac{\hat{Y}_{HT}}{\hat{X}_{HT}} X = \frac{\sum_{i\in s}\frac{y_i}{\pi_i}}{\sum_{i\in s}\frac{x_i}{\pi_i}} X, \tag{4.1}$$

where $\hat{Y}_{HT} = \sum_{i\in s}\frac{y_i}{\pi_i}$, $\hat{X}_{HT} = \sum_{i\in s}\frac{x_i}{\pi_i}$ and $X = \sum_{i\in U} x_i$.

From (4.1) we see that $\sum_{i\in s}\frac{y_i}{\pi_i}$ is a linear combination of study variable and sampling weight in selected sample. However, if some units of a study variable in a selected sample cannot be observed then $\hat{Y}_{HT}$ does not work. Therefore in this section we aim to propose a new ratio estimator for estimating population total in the presence of nonresponse. Recall from Definition 3.1, we introduced Srndal and Lundstrm [12]s estimator for Y under reverse framework and defined by $\hat{Y}_r^{(1)} = \sum_{i\in s}\frac{r_i y_i}{\pi_i p}$. Substitute $\hat{Y}_r^{(1)}$ instead of $\hat{Y}_{HT}$ in (4.1) we obtained the new ratio estimator denoted by $\hat{Y}_R^{(1)}$ and define in Definition 4.1.

Definition 4.1. Assuming that assumption (A) is correct under a reverse framework with unequal probability sampling without replacement. If p is known then the proposed estimator for Y is defined by,

$$\hat{Y}_R^{(1)} = \frac{\sum_{i\in s}\frac{r_i y_i}{\pi_i p}}{\sum_{i\in s}\frac{x_i}{\pi_i}} X, \tag{4.2}$$

where $X = \sum_{i\in U} x_i$.

Lemma 4.2. *Assuming that assumption (A) is correct under a reverse framework with unequal probability sampling without replacement. If p is unknown then the proposed estimator for Y is defined by $\hat{Y}_R^{(2)}$ and is given by,*

$$\hat{Y}_R^{(2)} = \frac{\sum_{i\in s}\frac{r_i y_i}{\pi_i \hat{p}}}{\sum_{i\in s}\frac{x_i}{\pi_i}} X = \frac{\sum_{i\in s}\frac{r_i y_i}{\pi_i}}{\sum_{i\in s}\frac{x_i}{\pi_i}} \frac{\sum_{i\in s}\frac{1}{\pi_i}}{\sum_{i\in s}\frac{r_i}{\pi_i}} X, \tag{4.3}$$

where X is defined in (4.2) and $\hat{p} = \Big(\sum\limits_{i\in s}\frac{r_i}{\pi_i}\Big)\Big(\sum\limits_{i\in s}\frac{1}{\pi_i}\Big)^{-1}$.

Proof. Under uniform nonresponse if p is unknown we can estimate p by $\hat{p} = \Big(\sum\limits_{i\in s}\frac{r_i}{\pi_i}\Big)\Big(\sum\limits_{i\in s}\frac{1}{\pi_i}\Big)^{-1}$ and substitute $\hat{p}$ instead of p in (4.2). □

Theorem 4.3. *Assuming that assumption (A) is correct under a reverse framework with unequal probability sampling without replacement. The proposed estimators are asymptotically unbiased estimators of Y.*

Proof. (i) Assuming that p is known the new ratio estimator for estimating population total is defined in (4.2). Under a reverse framework the expectation of $\hat{Y}_R^{(1)}$ is obtained by,

$$E(\hat{Y}_R^{(1)}) = E_R E_S(\hat{Y}_R^{(1)}|\mathbf{R}). \tag{4.4}$$

From (4.2) and (4.4) we see that $\hat{Y}_R^{(1)}$ has a form of nonlinear function. Therefore the linear function of $\hat{Y}_R^{(1)}$ can be obtained by using the Taylor linearization approach denoted by $\hat{Y}_{R,lin}^{(1)}$ and defined by,

$$\hat{Y}_{R,lin}^{(1)} \approx Y + (\hat{T}_1 - T_1) - \frac{T_1}{X}(\hat{T}_2 - T_2), \tag{4.5}$$

where $\hat{T}_1 = \sum_{i\in s}\frac{r_i y_i}{\pi_i p}, T_1 = \sum_{i\in U}\frac{r_i y_i}{p}, \hat{T}_2 = \sum_{i\in s}\frac{x_i}{\pi_i}, T_2 = X$. Substitute (4.5) in (4.4) we have,

$$E(\hat{Y}_R^{(1)}) \approx E_R E_S\Big[Y + (\hat{T}_1 - T_1) - \frac{T_1}{X}(\hat{T}_2 - T_2)\Big|\mathbf{R}\Big] = Y.$$

Therefore, $\hat{Y}_R^{(1)}$ is an asymptotically unbiased estimator of Y .

(ii) If p is unknown the proof is similar to (i).

□

4.2 The variance of the proposed estimator

In the presence of nonresponse, under a reverse framework with uniform nonresponse and an overall negligible sampling fraction Lawson and Ponkaew [13] discussed a method for investigated variance in estimation outcomes defined in Definition 4.4.

Definition 4.4. Under a reverse framework with an overall negligible sampling fraction let $\hat{Y}$ be the estimator of Y . The variance of $\hat{Y}$ is approximated by $V(\hat{Y}) \approx E_R V_S(\hat{Y}|\mathbf{R})$.

Next, we investigate variance of the proposed estimator. Recall from (4.2) if p is known we have,

$$\hat{Y}_R^{(1)} = \frac{\sum_{i\in s} \frac{r_i y_i}{\pi_i p}}{\sum_{i\in s} \frac{x_i}{\pi_i}} X = \hat{R}X, \tag{4.6}$$

where $\hat{R} = \Big(\sum_{i\in s} \frac{r_i y_i}{\pi_i p}\Big)\Big(\sum_{i\in s} \frac{x_i}{\pi_i}\Big)^{-1}$. Recall from Definition 4.4 and (4.6) the variance of $\hat{Y}_R^{(1)}$ is approximately,

$$V(\hat{Y}_R^{(1)}) \approx E_R V_S\big[Y_R^{(1)}|\mathbf{R}\big]. \tag{4.7}$$

Similar to Theorem 4.3 under Taylors linearization approach the linear function of $\hat{Y}_R^{(1)}$ is defined in (4.5) and we can rewrite $\hat{Y}_{R,lin}^{(1)}$ as,

$$\hat{Y}_{R,lin}^{(1)} \approx Constant + \sum_{i\in s} \frac{z_{1i}}{\pi_i}, \tag{4.8}$$

where $z_{1i} = \frac{r_i y_i}{\pi_i} - x_i R_r$ and $R_r = \frac{\sum_{i\in U} \frac{r_i y_i}{p}}{X}$.

Substitute $\hat{Y}_{R,lin}^{(1)}$ instead $\hat{Y}_R^{(1)}$ in (4.7) we have,

$$V(\hat{Y}_R^{(1)}) \approx E_R V_S\big[\sum_{i\in s} \frac{z_{1i}}{\pi_i}\Big|\mathbf{R}\big], \tag{4.9}$$

where z_{1i} is defined in (4.8).

Recall from (4.3) if p is unknown we have,

$$\hat{Y}_R^{(2)} = \frac{\sum_{i\in s} \frac{r_i y_i}{\pi_i \hat{p}}}{\sum_{i\in s} \frac{x_i}{\pi_i}} X = \frac{\sum_{i\in s} \frac{r_i y_i}{\pi_i}}{\sum_{i\in s} \frac{x_i}{\pi_i}} \frac{\sum_{i\in s} \frac{1}{\pi_i}}{\sum_{i\in s} \frac{r_i}{\pi_i}} X, \tag{4.10}$$

where X is defined in (4.2) and $\hat{p} = \Big(\sum\limits_{i\in s} \frac{r_i}{\pi_i}\Big)\Big(\sum\limits_{i\in s} \frac{1}{\pi_i}\Big)^{-1}$.

Recall from definition 4.4 and (4.10) the variance of $\hat{Y}_R^{(2)}$ is approximately,

$$V(\hat{Y}_R^{(2)}) \approx E_R V_S \big[\hat{Y}_R^{(2)} \Big| \mathbf{R}\big]. \tag{4.11}$$

Similar to equation (4.7) under Taylor linearization approach the linear function of $\hat{Y}_R^{(2)}$ is denoted by $\hat{Y}_{R,lin}^{(2)}$ and defined by,

$$\hat{Y}_{R,lin}^{(2)} \approx Constant + \sum_{i\in s} \frac{z_{2i}}{\pi_i}, \tag{4.12}$$

where $z_{2i} = \frac{N}{\sum_{i\in U} r_i}(y_i - \bar{Y}_R) - \frac{\bar{Y}_R}{\bar{X}}(x_i - \bar{X})$ and $\bar{Y}_R = \frac{\sum_{i\in U} r_i y_i}{\sum_{i\in U} r_i}$.

Substitute $\hat{Y}_{R,lin}^{(2)}$ instead of $\hat{Y}_R^{(2)}$ in (4.11) we have,

$$V(\hat{Y}_R^{(2)}) \approx E_R V_S \big[\sum_{i\in s} \frac{z_{2i}}{\pi_i} \Big| \mathbf{R}\big], \tag{4.13}$$

where z_{2i} defined in (4.12).

From (4.9) and (4.13) the general form of variance of the proposed ratio estimator may be written as,

$$V(\hat{Y}_R^{(m)}) \approx E_R V_S \big[\sum_{i\in s} \frac{z_{mi}}{\pi_i} \Big| \mathbf{R}\big], \tag{4.14}$$

where $m = 1, 2, z_{1i}$ is defined in (4.8) and z_{2i} id defined in (4.12).

Theorem 4.5. *Assuming that assumption (A) and (B) are correct under a reverse framework with unequal probability sampling without replacement.*

(i) *The variance of the proposed estimators can be derived from,*

$$V(\hat{Y}_R^{(m)}) \approx \sum_{i\in U} D_i E_R(z_{mi}^2) + \sum_{i\in U}\sum_{j\neq i\in U} D_{ij} E_R(z_{mi}) E_R(z_{mi}), \tag{4.15}$$

where $m = 1, 2, z_{1i}$ is defined in (4.8) and z_{2i} id defined in (4.12).

(ii) *If p is known then the variance of the proposed estimator is given by,*

$$V(\hat{Y}_R^{(1)}) \approx \sum_{i\in U} D_{2i} y_i^2 + \sum_{i\in U}\sum_{j\neq i\in U} D_{ij} e_i e_j - \sum_{i\in U} D_i (2y_i - Rx_i) Rx_i, \tag{4.16}$$

where $D_i = \frac{1-\pi_i}{\pi_i}, D_{2i} = \frac{1-\pi_i}{\pi_i p}, D_{ij} = \frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}$ and $e_i = y_i - Rx_i$.

(iii) *If p is unknown then the variance of the proposed estimator is given by,*

$$V(\hat{Y}_R^{(2)}) \approx \sum_{i\in U} D_{2i}(y_i - \bar{Y})^2 + \sum_{i\in U} D_i \tau + \sum_{i\in U}\sum_{j\neq i\in U} D_{ij} e_i e_j, \tag{4.17}$$

where D_{2i}, D_{ij} and e_i are defined in (4.16) and $\tau = \Big[(y_i - \bar{Y}) + (y_i - Rx_i)\Big](x_i - \bar{X})R$.

Proof. The proof of Theorem 4.5 is given in Appendix. □

5 Efficiency comparison

5.1 Theoretical analysis

In this section we compare the efficiency of the proposed estimator with Srndal and Lundstrm [12]'s estimator by comparing variance as discussed in Theorem 5.1.

Theorem 5.1. *Assuming that assumption (A) and (B) are correct under a reverse framework with unequal probability sampling without replacement. Let* $T_{1ij} = y_i x_j + y_j x_i - x_i x_j R$ *and* $T_{1i} = \Big[(y_i - \bar{Y}) + (y_i - Rx_i)\Big](x_i - \bar{X}) + 2(y_i - \bar{Y})\bar{Y}$.

(i) *Assume that p is known.* $V(Y_R^{(1)}) < V(\hat{Y}_r^{(1)})$ *if and only if* $\sum_{i\in U}\sum_{j\in U} D_{ij}T_{1ij} > 0$.

(ii) *Assume that p is unknown.* $V(Y_R^{(2)}) < V(\hat{Y}_r^{(2)})$ *if and only if* $\sum_{i\in U} D_i T_{1i} + \sum_{i\in U}\sum_{j\neq i\in U} D_{ij}T_{1ij} > 0$

Proof. (i) Let $T_{1ij} = y_i x_j + y_j x_i - x_i x_j R$ and assume that p is known.

$$
\begin{aligned}
V(\hat{Y}_R^{(1)}) &< V(\hat{Y}_r^{(1)}) \\
&\Leftrightarrow \sum_{i\in s} D_{2i} y_i^2 + \sum_{i\in U}\sum_{j\neq i\in U} D_{ij} e_i e_j - \sum_{i\in U} D_i (2y_i - Rx_i) R x_i \\
&< \sum_{i\in U} D_{2i} y_i^2 + \sum_{i\in U}\sum_{j\neq i\in U} D_{ij} y_i y_j \\
&\Leftrightarrow -\sum_{i\in s} D_i (2y_i x_i - Rx_i^2) R \\
&-\sum_{i\in U}\sum_{j\neq i\in U} D_{ij}(y_i x_j + y_j x_i - x_i x_j R) R < 0 \\
&\Leftrightarrow -\sum_{i\in U}\sum_{j\in U} D_{ij}(y_i x_j + y_j x_i - x_i x_j R) R < 0 \\
&\Leftrightarrow \sum_{i\in U}\sum_{j\in U} D_{ij}(y_i x_j + y_j x_i - x_i x_j R) R > 0 \\
&\Leftrightarrow \sum_{i\in U}\sum_{j\in U} D_{ij}(y_i x_j + y_j x_i - x_i x_j R) > 0 \\
&\Leftrightarrow \sum_{i\in U}\sum_{j\in U} D_{ij} T_{1ij} > 0. \qquad (5.1)
\end{aligned}
$$

Therefore when the condition (5.1) is satisfied, the proposed estimator is more efficient than Srndal and Lundstrm [12]s estimator.

(ii) The proof of (ii) is similar to the proof (i).

□

5.2 Simulation study

In this subsection, the efficiency of the proposed estimator is compared with the Srndal and Lundstrm [12]s estimator by using their simulated relative biases (RB) and relative root mean square errors ($RRMSE$) and defined by,

$$RB(\hat{Y}^{(m)}) = \frac{\frac{1}{B}\sum_{i=1}^{B}\hat{Y}_{[b]}^{(m)} - Y}{Y}, \tag{5.2}$$

$$RRMSE(\hat{Y}^{(m)}) = \frac{\sqrt{\frac{1}{B}\sum_{i=1}^{B}(\hat{Y}_{[b]}^{(m)} - Y)^2}}{Y}, \tag{5.3}$$

where $\hat{Y}^{(m)} = \hat{Y}_r^{(m)}$ or $\hat{Y}_R^{(m)}$. $\hat{Y}_r^{(m)}$ is the Srndal and Lundstrm [12]s estimator and $\hat{Y}_{r[b]}^{(m)}$ is the value of Srndal and Lundstrm [12]s estimator in iteration b, for $m = 1, 2$. $\hat{Y}_R^{(m)}$ is the proposed estimator and $\hat{Y}_{R[b]}^{(m)}$ is the value of proposed estimator in iteration b, for $m = 1, 2$. We use the Midzuno [15] scheme for selecting a sample s of size n from population U of size N. Under this scheme, the first and second order inclusion probabilities are given as follows,

$$\pi_i = \frac{k_i}{K}\frac{N-n}{N-1} + \frac{n-1}{N-1}, \tag{5.4}$$

$$\pi_{ij} = \frac{k_i + k_j}{K}\frac{N-n}{N-1}\frac{n-1}{N-1} + \frac{n-1}{N-1}\frac{n-2}{N-2}, \tag{5.5}$$

where $K = \sum_{i\in U} k_i$. The real population data of size $N = 299$ namely agstrat.dat was used in the simulation study (see, Lohr [16]). We used three variables from this data: the number of acres devoted to farms during 1992 (ACRES92), number of acres devoted to farms during 1982 (ACRES82) and number of farms (FARMS92). Let y = ACRES92, x =ACRES82, and k = FARMS92. The simulation steps to compare the efficiency of the Srndal and Lundstrm [12]s estimator with the proposed estimator are as follows.

Step1. Select a sample of 5 % from population U size $N = 299$.

Step 2. Generate a response indicator from the uniform nonresponse mechanism with different levels of response probability.

Step 3. Compute $\hat{Y}_r^{(m)}$ and $\hat{Y}_R^{(m)}$, where $m = 1, 2$.

Step 4. Repeat steps (1) to (3) 10,000 times.

Step 5. Compute $RB(\hat{Y}^{(m)})$ and $RRMSE(\hat{Y}^{(m)})$, where $\hat{Y}^{(m)} = \hat{Y}_r^{(m)}$ or $\hat{Y}_R^{(m)}$ and $m = 1, 2$.

Table 1 gives the empirical RB of the proposed estimator and Srndal and Lundstrm [12]s estimator for different response probability. The results show that for all levels of response probability the proposed estimator has a smaller value of RB than Srndal and Lundstrm [12]s estimator. Moreover, the RB of the proposed estimator and Srndal and Lundstrm [12]s estimator are very close to zero except where in the case that response probability is known and is less than 0.5. In addition, we present the simulated values of $RRMSE$ in table 2. The results show that the proposed estimator has a smaller value of $RRMSE$ than Srndal and Lundstrm [12]s estimator for all levels of response probability.

Table 1: The simulated RB of the Srndal and Lundstrm [12]s estimator and proposed estimator

p	Simulated relative biases			
	p is known		p is unknown	
	$RB(\hat{Y}_r^{(1)})$	$RB(\hat{Y}_R^{(1)})$	$RB(\hat{Y}_r^{(2)})$	$RB(\hat{Y}_R^{(2)})$
0.3	0.3412	0.3360	0.0087	0.0056
0.4	0.1141	0.1161	-0.0037	-0.0013
0.5	0.0086	0.0010	0.0051	-0.0020
0.6	0.0094	0.0041	0.0040	-0.0010
0.7	-0.0017	-0.0013	-0.0021	-0.0015
0.8	0.0040	-0.0011	0.0055	0.0002
0.9	-0.0027	-0.0005	-0.0031	-0.0007

Table 2: The simulated $RRMSE$ of the Srndal and Lundstrm [12]s estimator and proposed estimator

p	Simulated relative root mean square errors			
	p is known		p is unknown	
	$RRMSE(\hat{Y}_r^{(1)})$	$RRMSE(\hat{Y}_R^{(1)})$	$RRMSE(\hat{Y}_r^{(2)})$	$RRMSE(\hat{Y}_R^{(2)})$
0.3	0.7461	0.6011	0.4708	0.3401
0.4	0.5415	0.4177	0.4308	0.3083
0.5	0.4901	0.3653	0.4266	0.2904
0.6	0.4278	0.2916	0.3830	0.2290
0.7	0.3807	0.2398	0.3479	0.1877
0.8	0.3478	0.1869	0.3260	0.1448
0.9	0.3111	0.1243	0.3005	0.0983

6 Conclusions

In this paper we proposed a new ratio estimator for estimating population total using Srndal and Lundstrm [12]s estimator modified with Bacanli and Kadilar [5]s estimator with uniform nonresponse under unequal probability sampling without replacement. We investigated the variance of the proposed estimator under a reverse framework with an overall negligible sampling fraction. In theoretical analysis under some conditions we showed that the proposed estimator is more efficient than Srndal and Lundstrm [12]s estimator. The simulation study shows that the proposed estimator had RB and $RRMSE$ lower than Srndal and Lundstrm [12]s estimator for all levels of response probability.

Acknowledgement(s) : We are grateful to the referees for their helpful comments.

References

[1] W.G. Cochran, Sampling Techniques(3rd ed.) , New York: John Wiley & Sons (1977).

[2] B.V.S. Sisodia and V.K. Dwivedi, A modified ratio estimator using coefficient of variation of auxiliary variable, J. Indian Soc. Agric. Stat., 33 (1981) 13-18 .

[3] H.P. Singh and M.S. Kakran, A Modified Ratio Estimator Using Known Coefficient of Kurtosis of an Auxiliary Character, (1993). (Unpublished).

[4] L.N. Upadhyaya, and H.P. Singh , Use of transformed auxiliary variable in estimating the finite population mean, Biometrical J, 41(5) (1999) 627-636 .

[5] S. Bacanli, C. Kadilar , Ratio estimators with unequal probability designs, Pak. J. Statist, 24 (3) (2008)167-172 .

[6] B. J. Tepping, Variance Estimation in Complex Surveys. Proceedings of the . J. Am. Stat. Assoc, Social Statistics Section, 11-18 (1968).

[7] R.S. Woodruff , A Simple Method for Approximating the Variance of a Complicated Estimate. J. Am. Stat. Assoc. 66(334), 411-414 (1971).

[8] KM. Wolter, Introduction to Variance Estimation. New York: Springer-Verlag. (1985)

[9] A. Demnati, J. N. K. Rao, Linearization Variance Estimators for Survey Data with missing responses. Proceeding of the Section Survey Research Methods, J. Am. Stat. Assoc, 736-740 (2002).

[10] J.G. Bethlehem, Reduction of Nonresponse Bias through Regression Estimation. J. Offic. Stat, 4(3) 251-260 (1988).

[11] D.F. Horvitz, D.J. Thompson , A generalization of sampling without replacement from a finite universe. J. Am. Stat. Assoc, 47(260) 663-685 (1952).

[12] C.E. Srndal, S. Lundstrm, Estimation in Surveys with Nonresponse, New York: John Wiley & Sons (2005).

[13] N. Lawson, Variance estimation in the presence of nonresponse under probability proportional to size sampling. Proceedings of 6th Annual International Conference on Computational Mathematics, Computational Geometry & Statistics (CMCGS 2017), Singapore. (2017).

[14] N. Lawson, C. Ponkaew, New Generalized Regression Estimator in the Presence of Nonresponse under Unequal Probability Sampling. Comm Stat Theor Meth. (2018).

[15] H. Midzuno, On sampling system with probability proportional to sum of sizes, Ann. Inst.Statist. Math. 99-107 (1952).

[16] S.L. Lohr, . Sampling: Design and Analysis. Pacific Grove, CA: Duxbury Press. (1999).

Appendix:Proof of Theorme 4.5.

Assuming that (A) and (B) hold true.

Proof. (i) From (4.14) we have,

$$V(\hat{Y}_R^{(m)}) \approx E_R V_S\big[\sum_{i\in s}\frac{z_{mi}}{\pi_i}\Big|\mathbf{R}\big], \tag{A1}$$

where $z_{1i}=\frac{r_iy_i}{p}-x_iR_r$, $R_r=\frac{\sum_{i\in s}\frac{r_iy_i}{p}}{X}$ and $z_{2i}=\frac{N}{\sum_{i\in U}r_i}r_i(y_i-\bar{Y}_R)-\frac{\bar{Y}_R}{\bar{X}}(x_i-\bar{X})$. First of all we consider $V_S\Big[\sum_{i\in s}\frac{z_{mi}}{\pi_i}\Big|\mathbf{R}\Big]$ in (A1) since $\sum_{i\in s}\frac{z_{mi}}{\pi_i}$ has the form of Horvitz and Thompson [11]'s estimator therefore,

$$V_S\Big[\sum_{i\in s}\frac{z_{mi}}{\pi_i}\Big|\mathbf{R}\Big]=\sum_{i\in U}D_iz_{mi}^2+\sum_{i\in U}\sum_{j\neq i\in U}D_{ij}z_{mi}z_{mj}. \tag{A2}$$

Substituting (A2) into (A1) one has,

$$\begin{aligned}V(\hat{Y}_R^{(m)}) &\approx E_R\Big[\sum_{i\in U}D_iz_{mi}^2+\sum_{i\in U}\sum_{j\neq i\in U}D_{ij}z_{mi}z_{mj}\Big]\\ &=\sum_{i\in U}D_iE_R(z_{mi}^2)+\sum_{i\in U}\sum_{j\neq i\in U}D_{ij}E_R(z_{mi})E_R(z_{mj}).\end{aligned}$$

Therefore,

$$V(\hat{Y}_R^{(m)}) \approx \sum_{i\in U}D_iE_R(z_{mi}^2)+\sum_{i\in U}\sum_{j\neq i\in U}D_{ij}E_R(z_{mi})E_R(z_{mj}), \tag{A3}$$

where $z_{1i}=\frac{r_iy_i}{p}-x_iR_r$, $R_r=\frac{\sum_{i\in s}\frac{r_iy_i}{p}}{X}$ and $z_{2i}=\frac{N}{\sum_{i\in U}r_i}r_i(y_i-\bar{Y}_R)-\frac{\bar{Y}_R}{\bar{X}}(x_i-\bar{X})$.

(ii) From (A3) if $m = 1$ we have,

$$V(\hat{Y}_R^{(1)}) \approx \sum_{i \in U} D_i E_R(z_{1i}^2) + \sum_{i \in U} \sum_{j \neq i \in U} D_{ij} E_R(z_{1i}) E_R(z_{1j}). \tag{A4}$$

Next, we investigate $E_R(z_{1i})$ and $E_R(z_{1i}^2)$. Recall from (A3) we have $z_{1i} = \frac{r_i y_i}{p} - x_i R_r$ therefore,

$$E_R(z_{1i}) = E_R\Big[\frac{r_i y_i}{p} - x_i R_r\Big] = y_i - Rx_i = e_i, \tag{A5}$$

where $e_i = y_i - Rx_i$ and $R = \frac{\sum_{i \in U} y_i}{\sum_{i \in U} x_i}$.
Then,

$$z_{1i}^2 = \frac{r_i y_i^2}{p^2} - \frac{2 r_i y_i x_i R_r}{p} + x_i^2 R_r^2. \tag{A6}$$

From (A6) the expectation of z_{1i}^2 is obtained by,

$$E_R(z_{1i}^2) = E_R\Big[\frac{r_i y_i^2}{p^2} - \frac{2 r_i y_i x_i R_r}{p} + x_i^2 R_r^2\Big] \approx \frac{y_i^2}{p^2} - (2y_i - Rx_i) R x_i. \tag{A7}$$

Substitute (A5) and (A7) in (A4) we have,

$$V(\hat{Y}_R^{(1)}) \approx \sum_{i \in s} D_{2i} y_i^2 + \sum_{i \in U} \sum_{j \neq i \in U} D_{ij} e_i e_j - \sum_{i \in U} D_i (2y_i - Rx_i) R x_i, \tag{A7}$$

where $D_i = \frac{1-\pi_i}{\pi_i}$, $D_{2i} = \frac{1-\pi_i}{\pi_i p}$ and $D_{ij} = \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}$.

(iii) The proof of (iii) is similar to (ii).

□

(Received 27 August 2018)
(Accepted 30 December 2018)