

Research Article

Double Sampling with Ranked Set Selection in the Second Phase with Nonresponse: Analytical Results and Monte Carlo Experiences

Gaajendra K. Agarwal,¹ Sira M. Allende,² and Carlos N. Bouza²

¹ Software Development Division, Institute of Computing Training, Cuba

² Universidad de La Habana, Habana, Cuba

Correspondence should be addressed to Carlos N. Bouza, bouza@matcom.uh.cu

Received 19 May 2011; Revised 5 December 2011; Accepted 21 December 2011

Academic Editor: Man Lai Tang

Copyright © 2012 Gaajendra K. Agarwal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper is devoted to the study of the behavior of the use of double sampling for dealing with nonresponses, when ranked set sample is used. The characteristics of the sampling strategies are derived. The structure of the errors generated the need of studying of the optimality of the strategies by performing a set Monte Carlo experiments.

1. Introduction

The usual theory of survey sampling is developed assuming that the finite population $U = \{u_1, \dots, u_N\}$ is composed by individuals that can be perfectly identified. A sample s of size $n \leq N$ is selected. The variable of interest Y is measured in each selected unit. Real-life surveys should deal the existence of missing observations. There are three solutions to cope with this fact: to ignore the nonrespondents, to subsample the nonrespondents, or to impute the missing values. To ignore the non responses is a dangerous decision, to sub sample is a conservative and costly solution. Imputation is often used to compensate for item nonresponse. See, for discussions on the theme, Rueda and González [1], Singh [2], for example.

Section 2 presents the problem of non response when a single sample is selected.

We consider the use of double sampling for obtaining information on an auxiliary variable X . A first large sample is selected, it is supposedly noncostly. The values of X are used for selecting a ranked set sample (RSS), as the units are ranked using the values in the first stage sample. A selection of second sample provides a subsample from the preliminary large sample. The literature on the use of simple random double sampling (SRS) is large. Text books give the basic theory, see Singh [2] and Cochran [3]. In this paper we consider a ranked set

sampling (RSS) double sampling procedure. It is presented in Section 3 where a family of estimators is considered as an RSS alternative to the proposal of Singh and Kumar [4]. An expression of the gain in accuracy due to our proposed estimator is found. The estimator is compared with simple mean and the proposal of Singh and Kumar [4]. Real-life data are used for evaluating the behavior of these alternative estimators of the population mean in Section 4.

2. The Nonresponse Problem: A Single Sample

Non responses may be motivated by a refusal of some units to give the true value of Y or by other causes. Hansen and Hurvitz in 1946 [5] proposed selecting a sub-sample among the nonrespondents, see Cochran [3]. This feature depends heavily on the proposed sub-sampling rule. Sampling rules are due to Hansen and Hurvitz [5], Srinath [6], and Bouza [7]. The existence of non responses fixes that U is divided into two strata: $U_1 = \{u \in U \mid u \text{ responds at the first visit}\}$, $U_2 = U \setminus U_1$. Similarly s is partitioned into $s_i \subset U_i$, $i = 1, 2$. The procedure is a particular double sampling design described, using Hansen-Hurvitz's rule (HHR) as follows.

Step 1. Select a sample s from U using srswr.

Step 2. Evaluate Y among the respondents and determine $\{y_i : i \in s_1 \subset U_1, /s_1/ = n_1\}$. Compute

$$\bar{y}_1 = \frac{\sum_{i=1}^{n_1} y_i}{n_1}. \quad (2.1)$$

Step 3. Determine $n'_2 = n_2/K$, $K > 1$; $/s_2/ = n_2$ with $s_2 = \{u \in s \mid u \in U_2\}$.

Step 4. Select a sub-sample s'_2 of size n'_2 from s_2 using srswr.

Step 5. Evaluate Y among the units in $s'_2 \{y_i : i \in s'_2 \subset s_2, s_2 \subset U_2\}$. Compute

$$\bar{y}'_2 = \frac{\sum_{i=1}^{n'_2} y_i}{n'_2}. \quad (2.2)$$

Step 6. Compute the estimate of μ

$$\bar{\bar{y}} = \frac{n_1}{n} \bar{y}_1 + \frac{n_2}{n} \bar{y}'_2 = w_1 \bar{y}_1 + w_2 \bar{y}'_2. \quad (2.3)$$

Note that (2.1) is the mean of an srswr sample selected from U_1 , the response stratum, then its expected value is the mean of Y in the respondent stratum: μ_1 . We have that the conditional expectation of (2.2) is

$$E[\bar{y}'_2 \mid s] = \bar{y}_2, \quad (2.4)$$

as (2.4) is the mean of a srswr sample selected from the non response stratum U_2

$$EE[\bar{y}'_2 \mid s] = \mu_2, \quad (2.5)$$

and taking into account that for $i = 1, 2$ $E(n_i) = nN_i/N = nW_i$ the unbiasedness of (2.3) is easily derived.

The variance of (2.3) is deduced by using the following trick:

$$\bar{y} = (w_1\bar{y}_1 + w_2\bar{y}_2) + w_2(\bar{y}'_2 - \bar{y}_2), \quad (2.6)$$

the first term is the mean of s , then its variance is σ^2/n . For the second term we have that

$$\begin{aligned} V(w_2(\bar{y}'_2 - \bar{y}_2) | s) &= w_2^2 E(\bar{y}'_2 - \mu_2 - (\bar{y}_2 - \mu_2) | s)^2 \\ &= w_2^2 [E((\bar{y}'_2 - \mu_2) | s)^2 + E((\bar{y}_2 - \mu_2) | s)^2 \\ &\quad - 2E((\bar{y}'_2 - \mu_2)((\bar{y}_2 - \mu_2)) | s)]. \end{aligned} \quad (2.7)$$

Conditioning to a fixed n_2 we have that the expectation of the third term is $(\bar{y}_2 - \mu_2)^2$. Then we have that

$$\begin{aligned} V(w_2(\bar{y}'_2 - \bar{y}_2) | s) &= w_2^2 \left(\frac{\sigma_{2Y}^2}{n'_2} - \frac{\sigma_{2Y}^2}{n_2} \right) = w_2^2 \sigma_{2Y}^2 \left(\frac{K}{n_2} - \frac{1}{n_2} \right), \\ EV(w_2(\bar{y}'_2 - \bar{y}_2) | s) &= \frac{W_2(K-1)\sigma_{2Y}^2}{n}. \end{aligned} \quad (2.8)$$

Hence the expected error of (2.3) is given by the well-known expression

$$EV(\bar{y}) = \frac{\sigma_Y^2}{n} + \frac{W_2(K-1)\sigma_{2Y}^2}{n}. \quad (2.9)$$

Our proposal is to consider obtaining information provided by a known variable X for using RSS.

McIntire [8] proposed the method of RSS. He noticed the existence of a gain in accuracy with respect to the use of the sample mean with respect to srswr. Dell and Clutter [9] and Takahashi and Wakimoto [10] provided mathematical support to his claims. The following procedure provides a description of RSS selection.

2.1. RSS Procedure

Step 1. Randomly select m^2 units from the target population.

Step 2. Allocate the m^2 selected units as randomly as possible into m sets, each of size m .

Step 3. Without yet knowing any values for the variable of interest, rank the units within each set with respect to variable of interest. This may be based on personal professional judgment or done with concomitant variable correlated with the variable of interest.

Step 4. Choose a sample for actual quantification by including the smallest ranked unit in the first set, the second smallest ranked unit in the second set, the process is continued in this way until the largest ranked unit is selected from the last set.

Step 5. Repeat Steps 1 through 4 for r cycles to obtain a sample of size mr for actual quantification.

The RSS sample is the sequence of order statistics (OS) $\xi_{(1:1)t}, \dots, \xi_{(m:m)t}$, where $(j:h)t$ denotes the statistic of order j in the h th sample in the cycle $t = 1, \dots, r$. We have $n = mr$ observation and r of them are of the i th order statistics (os), $i = 1, \dots, m$. The RSS estimator of the mean of a variable of interest ξ , μ_{ξ} is

$$\mu_{(rss)\xi} = \frac{\sum_{t=1}^r \sum_{i=1}^m \xi_{(i:m)t}}{rm}, \quad (2.10)$$

and its variance is given by

$$V(\mu_{(rss)\xi}) = \frac{\sum_{i=1}^m \sigma_{\xi_{(i:m)}}^2}{rm^2} = \frac{\sigma_{\xi}^2}{rm} - \frac{\sum_{i=1}^m \Delta_{(i:m)}^2}{rm^2}, \quad (2.11)$$

where $\sigma_{\xi_{(i:m)}}^2 = E[\xi_{(i:m)} - E(\xi_{(i:m)})]^2$ and $\Delta_{(i:m)} = E(\xi_{(i:m)}) - \mu_{\xi}$.

The second term of (2.11) is the gain in accuracy due to the use of RSS instead of srswr.

Bouza [11] developed an RSS alternative under non responses. The non responses in s is $n_2 = rm_2$. He derived that, using a subsample size $m'_2 = m_2/K$,

$$\bar{y}'_{2rss} = \frac{\sum_{t=1}^r \sum_{i=1}^{m'_2} y_{(i:m'_2)t}}{rm'_2}, \quad (2.12)$$

is unbiased for the mean of Y in the nr stratum.

The cross-expectation's expected value is zero. In this case the RSS is balanced and we may express the variance of the order statistics (OS) as a function of the variance of Y in $U_2, V(y_{(i:m'_2)t})$, and the gains in accuracy measured by the $\Delta_{2Y(i)}^2$, 's as

$$V(\bar{y}_2 - \bar{y}'_{2rss} | s) = \sigma_{2Y}^2 \left(\frac{1}{n'_2} - \frac{1}{n_2} \right) - \sum_{i=1}^{m_2} \frac{\Delta_{2Y(i)}^2}{n'_2 m_2}. \quad (2.13)$$

Substituting $n'_2 = rm_2/K_2$ we obtain the following:

$$V(\bar{y}_{2rss} - \bar{y}'_{2rss} | s) = \frac{\sigma_{2Y}^2}{r} \left(\frac{K_2 - 1}{m_2} \right) - \sum_{i=1}^{m_2} \frac{\Delta_{2Y(i:m_2)}^2}{rm_2} \left(\frac{K_2 - 1}{m_2} \right) = V_2. \quad (2.14)$$

Taking the RSS estimator

$$\begin{aligned} \bar{\bar{y}}_{rss} &= \frac{n_1}{n} \bar{y}_{1rss} + \frac{n_2}{n} \bar{y}'_{2rss} = w_1 \bar{y}_{1rss} + w_2 \bar{y}'_{2rss}, \\ EV(\bar{\bar{y}}_{rss}) &= \frac{\sigma_Y^2}{n} + \frac{W_2(K-1)\sigma_{2Y}^2}{n} - \Psi(Y). \end{aligned} \quad (2.15)$$

Then there is gain in accuracy due to the use of RSS which is

$$\Psi(Y) = W_2(K-1)E\left(\frac{\sum_{i=1}^{m_2} \Delta_{2Y(i:m_2)}^2}{m_2}\right), \quad (2.16)$$

where $\Delta_{2Y(i:m)}^2 = (E(Y_{(i:m)} - \mu_Y)^2)$ is the gain in accuracy due to the use of RSS in the second stage.

3. The Nonresponse Problem: Double Sampling

We will consider that double sampling is used for obtaining a sample s^* from U using srswr. A cheap variable X is measured in the units in s^* . X is correlated with Y and we are able to compute the mean of it \bar{x} in the first stage. There are non responses. In the second stage we know $\bar{x}_{s^*} = (\sum_{i=1}^{n^*} x_i)/n^*$ and $\bar{\bar{x}} = (\sum_{i=1}^n x_i)/n$. Note that these estimates are used only in the estimation process.

Non responses on Y are present in the second stage sample and a subsample among the non respondents is selected. Singh and Kumar [4] considered this problem for simple random sampling. They proposed the family of estimators characterized by

$$\bar{y}^* = \bar{\bar{y}} \left(\frac{a\bar{\bar{x}} + b}{a\bar{x}_{s^*} + b} \right)^\alpha \left(\frac{a\bar{x} + b}{a\bar{x}_{s^*} + b} \right)^\beta, \quad \bar{\bar{y}} = \frac{\sum_{i=1}^n y_i}{n}. \quad (3.1)$$

The sampler fixes the constants α and β as well as a and b . They can be constants or functions, a different from zero. Taking

$$\varepsilon = \frac{\bar{\bar{y}} - \mu_Y}{\mu_Y}, \quad \theta = \frac{\bar{\bar{x}} - \mu_X}{\mu_X}, \quad \vartheta = \frac{\bar{x}_{s^*} - \mu_X}{\mu_X}, \quad \omega = \frac{\bar{x} - \mu_X}{\mu_X}. \quad (3.2)$$

Proposition 3.1 (see [4]). *The bias of*

$$\bar{y}^* = \bar{\bar{y}} \left(\frac{a\bar{\bar{x}} + b}{a\bar{x}_{s^*} + b} \right)^\alpha \left(\frac{a\bar{x} + b}{a\bar{x}_{s^*} + b} \right)^\beta \quad (3.3)$$

is

$$B(\bar{y}^*) = \mu_Y(\varphi_1 + \varphi_2), \quad (3.4)$$

defining

$$\begin{aligned} \varphi_1 &= \left(\gamma\phi \left[\alpha K_{xy} + \frac{\alpha-1}{2}\phi \right] + \beta \left(K_{xy} + \alpha\phi + \frac{\beta-1}{2}\phi \right) \right) c_x^2, \\ \varphi_2 &= \lambda\alpha\phi \left(K_{x_2y} + \frac{\alpha-1}{2}\phi \right) c_{x_2}^2, \end{aligned} \quad (3.5)$$

where

$$\begin{aligned} \gamma &= \frac{1}{n} - \frac{1}{n^*}, & \lambda &= \frac{W_2(K-1)}{n}, & c_x^2 &= \frac{\sigma_x^2}{\mu_x^2}, & c_{x_2}^2 &= \frac{\sigma_{x_2}^2}{\mu_{x_2}^2}, \\ K_{xy} &= \frac{\mu_x \sigma_{xy}}{\mu_y \sigma_x^2}, & K_{x_2y} &= \frac{\mu_{x_2} \sigma_{x_2y}}{\mu_y \sigma_{x_2}^2}, \\ \sigma_{xy} &= E(X - \mu_x)(Y - \mu_Y), & \sigma_{x_2y} &= E\left(\frac{(X - \mu_x)(Y - \mu_Y)}{U_2}\right). \end{aligned} \quad (3.6)$$

The variance is given by

$$V(\bar{y}^*) = \mu_Y^2 (\delta_1 + \delta_2), \quad (3.7)$$

defining

$$\begin{aligned} \delta_1 &= \left(\gamma \left[c_Y^2 + (\alpha + \beta) \phi((\alpha + \beta) \phi + 2K_{xy}) c_x^2 \right] \right), \\ \delta_2 &= \lambda \left(c_{y_2}^2 + \alpha \phi(\alpha \phi + 2K_{x_2y}) c_{x_2}^2 \right) + \frac{c_y^2}{n^*}, \\ c_y^2 &= \frac{\sigma_y^2}{\mu_y^2}, & c_{y_2}^2 &= \frac{\sigma_{y_2}^2}{\mu_{y_2}^2}. \end{aligned} \quad (3.8)$$

We are going to derive the RSS counterpart of this family. The first phase sample is selected using srswr and the information on X is used for selecting the initial sample and to subsample the non respondents. Our proposal is to use

$$\bar{y}_{\text{rss}}^* = \bar{\bar{y}}_{\text{rss}} \left(\frac{a\bar{\bar{x}}_{\text{rss}} + b}{a\bar{\bar{x}}_{s^*} + b} \right)^\alpha \left(\frac{a\bar{x} + b}{a\bar{x}_{s^*} + b} \right)^\beta, \quad (3.9)$$

$\bar{\bar{x}}_{\text{rss}}$ is the RSS mean of X in the second stage and

$$\varepsilon_{\text{rss}} = \frac{\bar{\bar{y}}_{\text{rss}} - \mu_Y}{\mu_Y}, \quad \theta_{\text{rss}} = \frac{\bar{\bar{x}}_{\text{rss}} - \mu_X}{\mu_X}, \quad \vartheta = \frac{\bar{x}_{s^*} - \mu_X}{\mu_X}, \quad \omega_{\text{rss}} = \frac{\bar{x}_{\text{rss}} - \mu_X}{\mu_X}. \quad (3.10)$$

Let us represent the involved estimators by

$$\begin{aligned} \bar{\bar{y}}_{\text{rss}} &= \mu_Y (1 + \varepsilon_{\text{rss}}), \\ \bar{\bar{x}}_{\text{rss}} &= \mu_X (1 + \theta_{\text{rss}}), \\ \bar{x}_{s^*} &= \mu_X (1 + \vartheta), \\ \bar{x}_{\text{rss}} &= \mu_X (1 + \omega_{\text{rss}}). \end{aligned} \quad (3.11)$$

Due to the unbiasedness of the estimators $E(X_{\text{rss}}) = 0$, $Z = \varepsilon, \theta, \vartheta, \omega$.

Taking

$$\phi = \frac{a\mu_X}{a\mu_X + b}. \quad (3.12)$$

We can rewrite (3.9) as

$$\bar{\bar{y}}_{\text{rss}}^* = \mu_Y \left[(1 + \varepsilon_{\text{rss}})(1 + \phi\theta_{\text{rss}})^\alpha (1 + \phi\vartheta)^{-\alpha} (1 + \phi\omega_{\text{rss}})^\beta (1 + \phi\vartheta)^{-\beta} \right]. \quad (3.13)$$

Note that

$$\begin{aligned} E(\varepsilon_{\text{rss}})^2 &= \frac{E(\bar{\bar{y}}_{\text{rss}} - \mu_Y)^2}{\mu_Y^2} = \frac{\sigma_Y^2/n + W_2(K-1)\sigma_{2Y}^2/n^*}{\mu_Y^2} \\ &\quad - \frac{W_2(K-1)E\left(\left(\sum_{i=1}^{m_2} \Delta_{2Y(i;m_2)}^2\right)/m_2\right)}{\mu_Y^2}, \\ E(\theta_{\text{rss}})^2 &= \frac{\sigma_X^2/n + W_2(K-1)\sigma_{2X}^2/n}{\mu_X^2} - \frac{W_2(K-1)E\left(\left(\sum_{i=1}^{m_2} \Delta_{2X(i;m_2)}^2\right)/nm_2\right)}{\mu_X^2}, \\ E(\vartheta)^2 &= \frac{E(\bar{x}_{s^*} - \mu_X)^2}{\mu_X^2} = \frac{\sigma_X^2}{n^*\mu_X^2}, \\ E(\omega_{\text{rss}})^2 &= \frac{\sigma_X^2/n - \left(\sum_{i=1}^m \Delta_{x(i)}^2\right)/rn}{\mu_X^2}. \end{aligned} \quad (3.14)$$

Under the hypothesis $|\phi Z| < 1$, $Z = \varepsilon_{\text{rss}}, \theta_{\text{rss}}, \vartheta, \omega_{\text{rss}}$, an expansion in Taylor series of (3.13) may be worked out. Grouping conveniently we have that

$$\begin{aligned} \bar{\bar{y}}_{\text{rss}}^* - \mu_Y &= \mu_Y \left[\varepsilon_{\text{rss}} + \beta(\omega_{\text{rss}} + \varepsilon_{\text{rss}}\omega_{\text{rss}} - \varepsilon_{\text{rss}}\vartheta) + \alpha\phi(\theta_{\text{rss}} + \varepsilon_{\text{rss}}\theta_{\text{rss}} - \varepsilon_{\text{rss}}\vartheta) \right. \\ &\quad \left. - (\alpha + \beta)\phi\vartheta + \alpha\beta\phi^2(\vartheta^2 + \vartheta(\omega_{\text{rss}} + \theta_{\text{rss}}) + \vartheta\omega_{\text{rss}}) - \phi^2(\beta^2\vartheta\omega_{\text{rss}} + \alpha^2\vartheta\theta_{\text{rss}}) \right. \\ &\quad \left. + \frac{\beta(\beta+1)\phi^2}{2}(\vartheta^2 + \omega_{\text{rss}}^2) + \frac{\alpha(\alpha+1)\phi^2}{2}(\vartheta^2 + \omega_{\text{rss}}^2) \right]. \end{aligned} \quad (3.15)$$

The cross-products for the OS $Z_{(i)}$, $Z = X, Y$, are expressed by

$$\begin{aligned} \sum_{i=1}^h (Z_{(i)} - \mu_{Z_{(i)}})(Z'_{(i)} - \mu_{Z'_{(i)}}) &= \sum_{i=1}^h (Z_{(i)} \mp \mu_Z - \mu_{Z_{(i)}})(Z'_{(i)} \mp \mu_{Z'} - \mu_{Z'_{(i)}}) \\ &= \sum_{i=1}^h (Z_{(i)} - \mu_Z)(Z'_{(i)} - \mu_{Z'}) - \sum_{i=1}^h Z_{(i)}\Delta_{Z'_{(i)}} + Z'_{(i)}\Delta_{Z_{(i)}} - \Delta_{Z_{(i)}}\Delta_{Z'_{(i)}} \\ &= (h-1)(\sigma_{ZZ'} + \Psi_{ZZ'}). \end{aligned} \quad (3.16)$$

The conditional expectations of the RSS estimators are

$$E(\bar{x}_{\text{rss}}/s^*) = E\left(E\left(\bar{x}_{\text{rss}}/s\right)/s^*\right) = \bar{x}^*. \quad (3.17)$$

Using these results we have that

$$\begin{aligned} E(\varepsilon_{\text{rss}}\theta_{\text{rss}}) &= \frac{\sigma_{XY} + \Psi_{XY}}{n\mu_x\mu_y} + \frac{W_2(K-1)(\sigma_{X_2Y} + \Psi_{X_2Y})}{n\mu_x\mu_y}, \\ E(\varepsilon_{\text{rss}}\vartheta) &= \frac{\sigma_{XY} + \Psi_{XY}}{n^*\mu_x\mu_y}, \\ E(\varepsilon_{\text{rss}}\omega_{\text{rss}}) &= \frac{\sigma_{XY} + \Psi_{XY}}{n\mu_x\mu_y}, \end{aligned} \quad (3.18)$$

with

$$\begin{aligned} \Psi_{X_2Y} &= -E\left(\frac{\sum_{i=1}^{m'_2} X_{(i)2}\Delta_{x_{(i)2}} + Y_{(i)2}\Delta_{y_{(i)2}} - \Delta_{x_{(i)2}}\Delta_{y_{(i)}}}{m_2}\right), \\ \Psi_{XY} &= -E\left(\frac{\sum_{i=1}^m X_{(i)}\Delta_{x_{(i)}} + Y_{(i)}\Delta_{y_{(i)}} - \Delta_{x_{(i)}}\Delta_{y_{(i)}}}{m}\right). \end{aligned} \quad (3.19)$$

In addition

$$\begin{aligned} E(\omega_{\text{rss}}\theta_{\text{rss}}) &= \frac{\sigma_x^2 + \Psi_X}{n\mu_x^2}, \quad \Psi_X = -\frac{\sum_{i=1}^m \Delta_{x_{8(i)}}^2}{r} \\ E(\vartheta\theta_{\text{rss}}) &= \frac{\sigma_x^2}{n^*\mu_x^2}, \\ E(\vartheta\omega_{\text{rss}}) &= \frac{\sigma_x^2}{n^*\mu_x^2}. \end{aligned} \quad (3.20)$$

Substituting in (3.15) after some algebraic work we obtain that the bias of (3.9) is

$$B(\bar{y}_{\text{rss}}^*) = \mu_Y(\varphi_{1\text{rss}} + \varphi_{2\text{rss}}), \quad (3.21)$$

where

$$\begin{aligned} \varphi_{1\text{rss}} &= \left(\gamma\phi\left[\alpha\left(K_{xy}c_x^2 + \frac{\Psi_{XY}}{n\mu_x\mu_y}\right) + \frac{\alpha-1}{2}\phi\left(c_x^2 + \frac{\Psi_X}{n\mu_x^2}\right)\right] \right. \\ &\quad \left. + \beta\left(K_{xy}c_x^2 + \frac{\Psi_{XY}}{n\mu_x\mu_y} + \alpha\phi\left(c_x^2 + \frac{\Psi_X}{n\mu_x^2}\right) + \frac{\beta-1}{2}\phi c_x^2\right) \right) \\ \Psi_{z_2} &= -\frac{E\left(\left(\sum_{i=1}^{m_2} \Delta_{2z(i;m_2)}^2\right)/m_2\right)}{n\mu_z^2}, \quad z = x, y. \end{aligned} \quad (3.22)$$

For a large value of n the bias tends to zero. Then we have proved the first statement of the following proposition.

Proposition 3.2. The estimator $\bar{y}_{rss}^* = \bar{y}_{rss}((a\bar{x}_{rss} + b)/(a\bar{x}_{s^*} + b))^\alpha ((a\bar{x}_{rss} + b)/(a\bar{x}_{s^*} + b))^\beta$ is asymptotically unbiased in terms of n and its variance is given by

$$V(\bar{y}_{rss}^*) = \frac{\sigma_Y^2}{n} + \gamma\mu_Y^2 \left(((\alpha + \beta)\phi)^2 c_x^2 + 2(\alpha + \beta)\phi K_{xy} c_x^2 + \frac{\Psi_{XY}}{\mu_x \mu_Y} \right) \\ + \lambda\mu_{Y_2}^2 \left(\frac{\sigma_{Y_2}^2}{\mu_{Y_2}^2} + \frac{\Psi_{Y_2}}{\mu_{Y_2}^2} + \alpha\phi \left(\alpha\phi \left(\frac{\sigma_x^2}{\mu_x^2} + \Psi_{x_2} \right) \right) + 2K_{x_2Y} c_{x_2}^2 + \frac{\Psi_{X_2Y}}{\mu_x \mu_Y} (1 + \Psi_{x_2}) + \frac{\sigma_{x_2Y}^2}{\mu_x \mu_Y} \right). \quad (3.23)$$

If $|\phi Z| < 1$, $Z = \varepsilon_{rss}, \theta_{rss}, \vartheta, \omega_{rss}$.

Proof. An expansion in Taylor series of $(\bar{y}_{rss}^* - \mu_Y)^2$ may be worked out. It is, neglecting the terms of order $t > 2$,

$$(\bar{y}_{rss}^* - \mu_Y)^2 = \mu_Y^2 (\tau_1 + \tau_2 + \tau_3 + \tau_4), \quad (3.24)$$

where

$$\begin{aligned} \tau_1 &= \varepsilon_{rss}^2 + (\alpha^2 \theta_{rss}^2 + \beta^2 \omega_{rss}^2 + 2\alpha\beta\varepsilon_{rss}\omega_{rss})\phi^2, \\ \tau_2 &= \varepsilon_{rss}^2 + (\alpha + \beta)^2 \vartheta^2 \phi^2, \\ \tau_3 &= 2\phi(\alpha\varepsilon_{rss}\theta_{rss} + \beta\varepsilon_{rss}\omega_{rss}), \\ \tau_4 &= -2(\alpha + \beta)(\phi\vartheta\varepsilon_{rss} + \phi^2(\alpha\vartheta\varepsilon_{rss} + \beta\vartheta\omega_{rss})). \end{aligned} \quad (3.25)$$

Calculating the expected value and grouping we have that

$$\begin{aligned} E(\bar{y}_{rss}^* - \mu_Y)^2 &= \frac{\sigma_Y^2}{n} + \gamma\mu_Y^2 \left(((\alpha + \beta)\phi)^2 c_x^2 + 2(\alpha + \beta)\phi K_{xy} c_x^2 + \frac{\Psi_{XY}}{\mu_x \mu_Y} \right) \\ &+ \lambda\mu_{Y_2}^2 \left(\frac{\sigma_{Y_2}^2}{\mu_{Y_2}^2} + \frac{\Psi_{Y_2}}{\mu_{Y_2}^2} + \alpha\phi \left(\alpha\phi \left(\frac{\sigma_x^2}{\mu_x^2} + \Psi_{x_2} \right) \right) \right. \\ &\quad \left. + 2K_{x_2Y} c_{x_2}^2 + \frac{\Psi_{X_2Y}}{\mu_x \mu_Y} (1 + \Psi_{x_2}) + \frac{\sigma_{x_2Y}^2}{\mu_x \mu_Y} \right). \end{aligned} \quad (3.26)$$

□

Remark 3.3. The gain in accuracy due to the use of (3.9) in terms of the variance is

$$G_{rss} = \frac{\sigma_{x_2y} + \gamma\mu_y^2 \Psi_{xy} + 2\Psi_{xy}(1 + \Psi_2) + \lambda\Psi_{x_2}\mu_y^2}{\mu_x \mu_y}. \quad (3.27)$$

Hence, as $V(\bar{y}_{rss}^*) = V(\bar{y}) + G$ the proposed method is more precise if $G < 0$.

This result allows to deduce the RSS counterparts of different double sampling estimators of the mean. For example,

$$\begin{aligned}
 (\alpha, \beta, a, b) &= (-1, 0, 1, 0) \longrightarrow \text{Khare-Srivanstava-Tabasum-Khan estimator 1,} \\
 (\alpha, \beta, a, b) &= (0, -1, 1, 0) \longrightarrow \text{Khare-Srivanstava-Tabasum-Khan estimator 2,} \\
 (\alpha, \beta, a, b) &= (-1, -1, 1, 0) \longrightarrow \text{Shing-Kumar ratio estimator,} \\
 (\alpha, \beta, a, b) &= (-1, 0, 1, 0) \longrightarrow \text{Shing-Kumar product estimator.}
 \end{aligned} \tag{3.28}$$

See Khare and Srivastava [12, 13] and Singh and Kumar, [4, 14, 15].

4. Numerical Comparisons

We compared the behavior of the proposed RSS method with the SRS one using data from three populations. Their description is given as follows.

Population 1

A set of 244 accounts was considered. The balance of each of them in the previous semester was X and Y was produced by an auditory. The first phase sample was provided by selecting 120 accounts and 72 non responses were reported. A new auditory was performed. The second stage sample was of size 24.

Population 2

The evaluation of radiographies provided values of X in 350 patients with cancer. A sample of 100 provided the first phase sample and 24 of them the second phase. Y was the size of an extirpated tumor. 53 measurements were missing. The measurement of them needed a search in the pathology department.

Population 3

The height of 1270 pigs provided the information on X in the population. 170 of them were selected at the first phase and 24 of them the second phase. Y was the weight of the pigs and 69 initial measurements were missing. The missing pig's weight was obtained by locating them before sending them to the butchery.

The values of r and m were fixed conveniently for obtaining a sample of size 24. The means and variances of the os's involved were determined by forming all the possible samples and computing them. The relative gain in accuracy due to the use of RSS was measured by

$$\varpi = \frac{G_{\text{RSS}}}{V(\bar{y}^*)}, \tag{4.1}$$

Table 1: Gain in accuracy due to the use of RSS in three populations.

Population	$m = 3$	$m = 4$	$m = 6$
Balance of accounts	0,1122	0,1523	0,1095
Size of tumors	0,1214	0,1207	0,1105
Height of pigs	0,2672	0,2998	0,2159

Table 2: Gain in accuracy due to the use of RSS of six populations: $n^* = 240$ and $K = 0, 10$.

Distribution	$m = 3$	$m = 4$	$m = 6$
Uniform (0,1)	0,121	0,146	0,118
Normal (0,1)	0,101	0,127	0,096
Logistic (0,1)	0,009	0,011	0,009
Laplace (0,1)	0,087	0,092	0,074
Exponential (1)	0,006	0,007	0,005
Gamma (2,1)	0,092	0,114	0,088
Weibull (1,3)	0,081	0,087	0,074
Beta (7,4)	0,157	0,151	0,138

for $m = 3, 4, 6$. The results are given in Table 1. They sustain that the use of RSS provides gains of accuracy larger than 10%.

A similar study was developed by generating a sample of 240 values of X and determining

$$Y = 5 + 2X + \varepsilon, \quad (4.2)$$

ε was generated using the same distribution. The results are given in Table 2. Note that generally the gain in efficiency is larger when the underlying distribution is symmetric. The best results are derived when $m = 4$ excepting the Beta distribution.

5. Conclusions

The accuracy of the proposed method seems to be better than the SRS method when G_{RSS} is analyzed. It can take negative values but it has been larger than zero in the experiments developed. It was around 0,1 in all the cases and using $m = 4$ may be the best choice.

Acknowledgments

The authors thank the referees for their helpful comments which allowed improving a previous version. This paper was supported by the CONACYT Contract 10110/62/10, FON. INST. 8/10.

References

- [1] M. Rueda and S. González, "Missing data and auxiliary information in surveys," *Computational Statistics*, vol. 19, no. 4, pp. 551–567, 2004.

- [2] S. Singh, *Advanced Sampling Theory with Applications*, Kluwer Academic, Dordrecht, The Netherlands, 2003.
- [3] W. G. Cochran, *Sampling Techniques*, Wiley and Sons, New York, NY, USA, 1971.
- [4] H. P. Singh and S. Kumar, "A general procedure of estimating the population mean in the presence of non-response under double sampling using auxiliary information," *Statistics and Operations Research Transactions*, vol. 33, no. 1, pp. 71–84, 2009.
- [5] M. H. Hansen and W. N. Hurvitz, "The problem of non responses in survey sampling," *Journal of American Statistical Association*, vol. 41, pp. 517–523, 1946.
- [6] K. P. Srinath, "Multiphase sampling in nonresponse problems," *Journal of the American Statistical Association*, vol. 66, pp. 583–589, 1971.
- [7] C. N. Bouza, "Sobre el problema de la fraccion de submuestreo para el caso de las no respuestas," *Trabajos de Estadística y de Investigacion Operativa*, vol. 32, no. 2, pp. 30–36, 1981.
- [8] G. A. McIntire, "A method for unbiased sampling using ranked sets," *Australian Journal of Agricultural Research*, vol. 3, pp. 385–390, 1952.
- [9] T. R. Dell and J. L. Clutter, "Ranked set sampling theory with order statistics background," *Biometrics*, vol. 28, pp. 545–555, 1972.
- [10] K. Takahashi and M. Wakimoto, "On unbiased estimates of the population mean based on the sample stratified by means ordering," *Annals of the Institute of Mathematical Statistics*, vol. 20, no. 1, pp. 1–31, 1967.
- [11] C. N. Bouza, "Estimation of the mean in ranked set sampling with non responses," *Metrika*, vol. 56, no. 2, pp. 171–179, 2002.
- [12] B. B. Khare and S. Srivastava, "Estimation of population mean using auxiliary character in presence of non-response," *National Academy of Science Letters*, vol. 16, pp. 111–114, 1993.
- [13] B. B. Khare and S. Srivastava, "Study of conventional and alternative two phase sampling ratio product and regression estimators in presence of non-response," *Proceedings of the National Academy of Sciences*, vol. 65, pp. 195–203, 1995.
- [14] H. P. Singh and S. Kumar, "Estimation of mean in presence of non-response using two phase sampling scheme," *Statistical Papers*, vol. 51, no. 3, pp. 559–582, 2010.
- [15] H. P. Singh and S. Kumar, "A regression approach to the estimation of the finite population mean in the presence of non-response," *Australian & New Zealand Journal of Statistics*, vol. 50, no. 4, pp. 395–408, 2008.

