STATISTICA, anno LXXIV, n. 3, 2014

REPLICATION VARIANCE ESTIMATION UNDER TWO-PHASE SAMPLING IN THE PRESENCE OF NON-RESPONSE

Muqaddas Javed[1]
National College of Business Administration and Economics, Lahore, Pakistan.
Zahoor Ahmad
Department of Statistics, University of Gujrat, Gujrat, Pakistan.
Muhammad Hanif
National College of Business Administration and Economics, Lahore, Pakistan.

1. Introduction

Two phase sampling introduced by Neyman (1938) he give the concept of use of auxiliary information to get maximum response about the required information. He also gave the idea of the use of stratification in the second phase. Hansan and Hurwitz (1946) suggest the survey statistician to get maximum response by re-contacting with expansive method to non respondent. A simple method of double sampling for stratification is proposed and the classical non-response theory is obtained as a special case by Rao (1973). Ratio and regression estimator of the population mean was proposed by Cochran (1977) for the study variable in which information on the auxiliary variable is obtained from all the sample units. More improvement in population mean in the presence of non response using auxiliary information was suggested by Rao (1986). Singh et al. (1990) proposed a class of estimators based on general sampling designs for population parameter utilizing auxiliary information of some other parameters. They also studied the general properties for the suggested class and find the asymptotic lower bound to the mean square error of the estimators belonging to the class. They (1990) also proposed several unbiased ratio and product estimators with their expressions of asymptotic variances using Jackknife technique in two-phase sampling.

Double expansion estimator (DEE) and reweighted expansion estimator (REE) are developed by Kott and Stukel (1997) for regression estimator in two phase stratified sampling. Khare and Srivastve (1997) proposed some improved estimator of Ratio and Regression using auxiliary information in the presence of non response.Precise variance estimator is always a choice of survey statistician; replication variance estimation is a tool which overcomes this problem. The concept of the jackknife was introduced by Quenouille (1956) in connection with reduction

[1] Corresponding Author. E-mail: muqaddas.javed@uog.edu.pk

of bias for nonlinear estimation. The possibility of using this technique for the purpose of variance estimation was proposed by Tukey (1958). He (1958) suggested that each jackknife replicate estimate might be regarded as an independent and identically distributed random variable, which in turn suggests a very simple variance estimator. Jones (1958) modified the jackknife variance estimator for the population mean under two phase sampling for ratio estimator. Durbin (1959) may have been the first to use it in the context of finite populations.

Rao and Shao (1992) proposed a consistent jackknife variance estimator for reweighted expansion estimator in the context of hot deck imputation. Breidt and Fuller (1993) suggested a replication variance estimator for multiphase samples. Their (1993) procedure is particularly applicable if the first phase and second phase primary sampling units are the same and if there are at least two second phase primary sampling units in each first phase stratum. Rust and Rao (1995) provide comprehensive overviews on replication variance estimation topic. A jackknife procedure has been suggested for the ratio estimator by Rao and Sitter (1995) and for the regression estimator by Sitter (1997). Kott (1990) discuss variance estimation in a situation where the stratified design for the second phase differs from the stratified design for the first phase, and suggests a jackknife estimator of variance for a particular two phase estimator. Kott and Stukel (1997) concluded that the jackknife variance estimator works well for the reweighted expansion estimator if the first phase sampling rate is negligible.Fuller (1998) give the estimation of the variance of regression estimator for a two phase sample and replication variance estimator is developed in second phase sample. Kim and Sitter (2003) developed a replication variance estimator for two phase sampling when the second phase sample is much smaller than the first phase sample; Method is applied to DEE and REE. A consistent variance estimator which is applicable in DEE and REE in two phase sampling was developed by Kim, Navarro and Fuller (2006) when the second phase is stratified. Wolter (2007) provide comprehensive overviews on replication variance estimation topic. Ramasubramanian et al. (2007) proposed a two new jackknife method as the counterparts of two existing Bootstrap methods of variance estimation under two phase sampling.Farrell and Singh (2010) discuss the problem of estimating the variance of various estimators of the population mean in two phase sampling by jackknifing the two phase calibrated weights. They (2010) also estimated the variance of the chain ratio and regression type estimators due to Chand (1975) using the jackknife. Kim and Yu (2011) developed a replication based bias adjusted variance estimator that extends the method of Kim, Navarro and Fuller (2006) under two phase sampling.

In case of non-response the sampler has an incomplete sample data that affects the quality of estimates of the unknown population parameters. The problem was first considered by Hansen and Hurwitz (1946), in which the population was divided into groups of respondents and non-respondents and the sample size to be attempted at the first occasion and the proportion of non-respondents to be repeated in the remaining sample was also determined. They constructed the estimate of the population mean by combining the data from the two attempts and derived the expression for the sampling variance of the estimates and optimum sampling fraction among the non respondents is also obtained.

In this paper, we propose some new replication variance estimators in the presence of non-response under two phase sampling. The estimators are proposed in two different situations, Kim and Yu (2011) methodology is used for the replication.

The rest of the paper is organized as follows. In section 2, the four estimators are developed under two phase sampling in the presence of non response and their replication variance estimators are developed in the Section 3. In Section 4, proposed method is extended for DEE-type estimators and REE-type estimators. In section 5, a limited simulation study is presented. Concluding remarks are made in last section of the paper.

2. Proposed Estimators of Mean under Two-phase Sampling in the Presence of Non-response

In this section, we suggest two phase stratified estimators for estimating population mean of variable of interest in the presence of non response. The estimators are proposed for two different conditions.

(i) Estimation of mean in the presence of non-response at first phase

(ii) Estimation of mean in the presence of non-response at second phase

2.1. Estimation of Mean in the Presence of Non-response at First Phase Sample

Here, we suggest two phase stratified estimators for estimating population mean of the study variable when non-response is present on first phase under the following two different situations.

(i) Estimation by ignoring non-responding units

(ii) Estimation by re-contacting non responding units

2.1.1. Estimation by Ignoring Non-responding Units

In this section, we simply assume the situation where the first phase sample of size n is selected with simple random sampling without replacement (SRSWOR) from a finite population of size N and selected first phase sample is divided into two parts, respondent n_r and non-respondent $n_{\bar{r}}$. Using the information obtained from first phase responding sample n_r, it is stratified into L strata for second phase sampling. We have n_{rh} first phase responding sample elements and assume A_{rh1} be the set of indices for the first phase responding sample elements in stratum h. A stratified random sample of size $n'$with sample size n'_{rh} in stratum h is selected in the second phase sampling. Where $n' = \sum_{h=1}^{L} n'_{rh}$ and n_{rh}/n is sampling rate for each stratum.

The study variable y_i is observed in the second phase sample; the population means of y is estimated by

$$t_1 = \frac{1}{n_r}\sum_{h=1}^{L}\sum_{i\in A_{rh2}}\frac{n_{rh}}{n'_{rh}}y_{rhi}, \quad (1)$$

where A_{rh2} is the set of indices for the second phase sample elements that belong to stratum h. Using proof 1 of Appendix, the variance of t_1 can be written as

$$V(t_1) = E\Big[n_r^{-1}(1-f_1)\sum_{h=1}^{L}w_h\left(\bar{y}_{rh1}-\bar{y}_1\right)^2 + \sum_{h=1}^{L}\left({n'}_{rh}^{-1} - n_{rh}^{-1}f_1\right)w_h^2 s_{rh1}^2\Big], \quad (2)$$

where $\bar{y}_1 = \sum_{i=1}^{n_r} y_{ri}/n_r$ be the mean of the first phase sample and the consistent estimator of the variance of t_1 can be derived by replacing $\bar{y}_{h1}\, and\, s_{h1}^2$ by their estimates $\bar{y}_{rh2} = {n'}_{rh}^{-1}\sum_{i\in A_{rh2}} y_{ri}$ and $s_{rh2}^2 = \left({n'}_{rh}-1\right)^{-1}\sum_{i\in A_{rh2}}\left(y_{ri}-\bar{y}_{rh2}\right)^2$ respectively in (2). So the consistent estimator of (1) is

$$\hat{V}(t_1) = n_r^{-1}(1-f_1)\sum_{h=1}^{L}w_h\left(\bar{y}_{rh2}-\bar{y}_2\right)^2 + \sum_{h=1}^{L}\left({n'}_{rh}^{-1} - n_{rh}^{-1}f_1\right)w_h^2 s_{rh2}^2, \quad (3)$$

where $f_1 = \frac{n_r}{N}$, $\bar{y}_2 = \sum_{h=1}^{L} w_h\bar{y}_{h2}$ and $w_h = \frac{n_{rh}}{n_r}$.

2.1.2. Estimation by Re-contacting non Responding Units

Here, we develop an estimator by the theory of Hansan and Hurwitz (1946) to re-contact the non-responding units with a different method. In this situation the first phase sample units is stratified for both responding units n_r and non-responding units $n_{\bar{r}}$ into L strata for second phase sampling.We have n_{rh} first phase responding sample elements and assume A_{rh1} be the set of indices for the first phase responding sample elements in stratum h and we also have $n_{\bar{r}h}$ first phase non-responding sample elements in stratum h. In the second phase sampling, a stratified random sample of size n'_1 is selected from responding units with sample size n'_{1h} in stratum h, and a stratified random sample of size n'_2 is selected from re-contacted first phase non-responding units with sample size n'_{2h} in stratum h, where $n'_1 = \sum_{h=1}^{L} n'_{1\mathrm{h}}$, $n'_2 = \sum_{h=1}^{L} n'_{2h}$ and a fixed sampling rate for each first phase responding stratum at second phase is n'_{1h}/n_{rh} and first phase non-responding stratum at second phase is $n'_{2h}/n_{\bar{r}h}$.

The study variable y_i is observed in the second phase sample, the population mean of y is estimated by

$$t_2 = \frac{w_1}{n_r}\sum_{h=1}^{L}\sum_{i\in A_{rh2}}\frac{n_{rh}}{n'_{1h}}y_{rhi} + \frac{w_2}{n_{\bar{r}}}\sum_{h=1}^{L}\sum_{i\in A_{\bar{r}h2}}\frac{n_{\bar{r}h}}{n'_{2h}}y_{\bar{r}hi}, \quad (4)$$

where $w_1 = \frac{n'_1}{n_r}$, $w_2 = \frac{n'_2}{n_{\bar{r}}}$ and A_{rh2} is the set of indices for the second phase responding sample elements that belong to stratum h, $A_{\bar{r}h2}$ is the set of indices for the second phase re-contacting sample elements that belong to stratum h. Using the result given in the proof 2 of Appendix, the variance of estimator t_2 can be written as

$$\hat{V}(t_2) = \frac{n'_1(1-f_1)}{n_r^{-1}} \sum_{h=1}^{L} w_h (\bar{y}_{rh2} - \bar{y}_{r2})^2 + w_1^2 \sum_{h=1}^{L} \left(\frac{1}{n'^{-1}_{1h}} - \frac{f_1}{n_{rh}^{-1}} \right) w_h^2 s_{rh2}^2$$

$$+ \frac{n'_2(1-f_2)}{n_{\bar{r}}^{-1}} \sum_{h=1}^{L} w_h^* (\bar{y}_{\bar{r}h2} - \bar{y}_{\bar{r}2})^2 + w_2^2 \sum_{h=1}^{L} \left(\frac{1}{n'^{-1}_{2h}} - \frac{f_2}{n_{\bar{r}h}^{-1}} \right) w_h^{*2} s_{\bar{r}h2}^2, \quad (5)$$

2.2. *Estimation of Mean in the Presence of Non-response at Second Phase Sample*

In this section we again suggest two new estimators for the population mean in two phase stratified design when non-response is present on second phase under the following situations.

(i) Estimation by ignoring non-responding units

(ii) Estimation by re-contacting non responding units

2.2.1. *Estimation by Ignoring Non-responding Units*

In this section, we assume the situation where the first phase sample of size n is selected with SRSWOR from a finite population of size N. Using the information obtained from first phase sample, it is stratified into L strata for second phase sampling. We have n_h first phase sample elements and assume A_{h1} be the set of indices for the first phase sample elements in stratum h. A stratified random sample of size n' with sample size n'_h in stratum h is selected in the second phase sampling of which n'_{rh} respond and $n'_{\bar{r}h}$ non-respondent. Where $n' = \sum_{h=1}^{L} n'_h = \sum_{h=1}^{L} (n'_{rh} + n'_{\bar{r}h})$ and a fixed sampling rate for each stratum is n'_{rh}/n_h.

The study variable y_i is observed in the second phase sample; the populations mean of y is estimated by

$$t_3 = \frac{1}{n} \sum_{h=1}^{L} \sum_{i \in A_{rh2}} \frac{n_h}{n'_{rh}} y_{hi}, \quad (6)$$

where A_{rh2} is the set of indices for the second phase sample elements that belong to stratum h. The variance of t_3 can be written as

$$V(t_3) = E\Big[n^{-1}(1-f_1^*) \sum_{h=1}^{L} w_h^\circ (\bar{y}_{rh1} - \bar{y}_1)^2 + \sum_{h=1}^{L} \left(n'^{-1}_{rh} - n_h^{-1} f_1^* \right) w_h^{\circ 2} s_{rh1}^2 \Big]. \quad (7)$$

A consistent estimator of the variance of t_3 can be derived by replacing $\bar{y}_{h1}$ *and* s_{h1}^2 with their estimates $\bar{y}_{rh2} = n'^{-1}_{rh} \sum_{i \in A_{rh2}} y_{ri}$ and

$s^2_{rh2} = \left(n'_{rh} - 1\right)^{-1} \sum_{i \in A_{rh2}} \left(y_{ri} - \bar{y}_{rh2}\right)^2$ respectively in (7). So the consistent estimator of (6) is

$$\hat{V}\left(t_3\right) = n^{-1}\left(1 - f_1^*\right) \sum_{h=1}^{L} w_h^{\circ} \left(\bar{y}_{rh2} - \bar{y}_2\right)^2 + \sum_{h=1}^{L} \left(n'^{-1}_{rh} - n_h^{-1} f_1^*\right) w_h^{\circ 2} s^2_{rh2}, \quad (8)$$

where $f_1^* = \frac{n}{N}$, $\bar{y}_2 = \sum_{h=1}^{L} w_h \bar{y}_{rh2}$ and $w_h^{\circ} = \frac{n_h}{n}$.

2.2.2. *Estimation by re-contacting non responding units*

In this section, we assume the situation where the first phase sample of size n is selected with SRSWOR from a finite population of size N. Using the information obtained from first phase sample, it is stratified into L strata for second phase sampling. We have n_h first phase sample elements and assume A_{h1} be the set of indices for the first phase responding sample elements in stratum h. A stratified random sample of size n' with sample size n'_h in stratum h is selected in the second phase sampling of which n'_{rh} respond and $n'_{\bar{r}h}$ non-respondent. Where $n' = \sum_{h=1}^{L} n'_h = \sum_{h=1}^{L} \left(n'_{rh} + n'_{\bar{r}h}\right)$ and a fixed sampling rate for each stratum responding unit is n'_{rh}/n_h and non-responding unit is $n'_{\bar{r}h}/n_h$.

The study variable y_i is observed in the second phase sample, the population mean of y is estimated by

$$t_4 = \frac{1}{n} \sum_{h=1}^{L} \sum_{i \in A_{rh2}} w^*_{h1} \frac{n_h}{n'_{rh}} y_{rhi} + \frac{1}{n} \sum_{h=1}^{L} \sum_{i \in A_{\bar{r}h2}} w^*_{h1} \frac{n_h}{n'_{\bar{r}h}} y_{\bar{r}hi}, \quad (9)$$

where $w^*_{h1} = \frac{n_h}{n'}$ and A_{rh2} is the set of indices for the second phase responding sample elements that belong to stratum h and $A_{\bar{r}h2}$ is the set of indices for the second phase re-contacting sample elements that belong to stratum h. The variance of t_4 can be written as

$$\begin{aligned} \hat{V}(t_4) = & \sum_{h=1}^{L} \left[\left(w' - w^{*2}_{h1}\right) n_h^{-1} + \left(w^{*2}_{h1} n'^{-1}_{rh} - n_h^{-1} w' f_1^*\right)\right] w_h^{\circ 2} s^2_{rh2} \\ & + \sum_{h=1}^{L} \left[\left(w' - w^{*2}_{h1}\right) n_h^{-1} + \left(w^{*2}_{h1} n'^{-1}_{\bar{r}h} - n_h^{-1} w' f_1^*\right)\right] w_h^{\circ 2} s^2_{\bar{r}h2} \quad (10) \\ & + w' n^{-1} \left(1 - f_1^*\right) \sum_{h=1}^{L} w_h^{\circ} \left[\left(\bar{y}_{rh2} - \bar{y}_2\right)^2 + \left(\bar{y}_{\bar{r}h2} - \bar{y}_2\right)^2\right], \end{aligned}$$

where $f_1^* = \frac{n}{N}$, $\bar{y}_2 = \sum_{h=1}^{L} w_h \bar{y}_{rh2}$, $w' = \frac{n^2}{(n')^2}$, $\bar{y}_{rh2} = n'^{-1}_{rh} \sum_{i \in A_{rh2}} y_{ri}$, $w_h^{\circ} = \frac{n_h}{n}$, $\bar{y}_{\bar{r}h2} = n'^{-1}_{\bar{r}h} \sum_{i \in A_{\bar{r}h2}} y_{\bar{r}i}$.

3. Replication Variance Estimation of Proposed Estimators

In this section we proposed some replication variance estimators for the estimators developed in section 2, the proposed method of jackknife replication variance estimation of Kim and Yu (2011) by successively deleting units from the entire first phase sample is adopted. In this method first phase sampling rate is not necessarily negligible and does not require additional replicates for bias correction in the variance estimation.
The full jackknife variance estimator of (1) is

$$\hat{V}_{J1} = \sum_{k \in A_1} c_k (t_1^{(k)} - t_1)^2, \tag{11}$$

where $t_1^{(k)} = \frac{1}{n_r^*} \sum_{h=1}^{L} n_{rh}^* \bar{y}_{rh2}^{(k)}$ and k is the index of the units deleted in the jackknife replicates, A_1 is the set of indices for the first phase sample elements and c_k is a factor associated with replicate k determined by the replication method

$$\frac{1}{n_r^*} n_{rh}^* = \begin{cases} (n_r - 1)^{-1}(n_{rh} - 1) & if \;\; k \in A_{rh1} \\ (n_r - 1)^{-1} n_{rh} & if \;\; k \notin A_{rh1} \end{cases}$$

and

$$\bar{y}_{rh2}^{(k)} = \begin{cases} (n'_h - 1)^{-1}(n'_h \bar{y}_{rh2} - y_{rk}) & if \;\; k \in A_{rh2} \\ \bar{y}_{rh2} & if \;\; k \notin A_{rh2}. \end{cases}$$

The full jackknife variance estimator of (4) is

$$\hat{V}_{J2} = \sum_{k \in A_1} c_k (t_2^{(k)} - t_2)^2, \tag{12}$$

where $t_2^{(k)} = \frac{w_1}{n_r^*} \sum_{h=1}^{L} n_{rh}^* \bar{y}_{rh2}^{(k)} + \frac{w_2}{n_{\bar{r}}^*} \sum_{h=1}^{L} n_{\bar{r}h}^* \bar{y}_{\bar{r}h2}^{(k)}$
where

$$\frac{1}{n_r^*} n_{rh}^* = \begin{cases} (n_r - 1)^{-1}(n_{rh} - 1) & if \;\; k \in A_{rh1} \\ (n_r - 1)^{-1} n_{rh} & if \;\; k \notin A_{rh1}, \end{cases}$$

$$\frac{1}{n_{\bar{r}}^*} n_{\bar{r}h}^* = \begin{cases} (n_{\bar{r}} - 1)^{-1}(n_{\bar{r}h} - 1) & if \;\; k \in A_{\bar{r}h1} \\ (n_{\bar{r}} - 1)^{-1} n_{\bar{r}h} & if \;\; k \notin A_{\bar{r}h1}, \end{cases}$$

$$\bar{y}_{rh2}^{(k)} = \begin{cases} (n'_{1h} - 1)^{-1}(n'_{1h} \bar{y}_{rh2} - y_{rk}) & if \;\; k \in A_{rh2} \\ \bar{y}_{rh2} & if \;\; k \notin A_{rh2} \end{cases}$$

and

$$\bar{y}_{\bar{r}h2}^{(k)} = \begin{cases} (n'_{2h} - 1)^{-1}(n'_{2h} \bar{y}_{\bar{r}h2} - y_{\bar{r}k}) & if \;\; k \in A_{\bar{r}h2} \\ \bar{y}_{\bar{r}h2} & if \;\; k \notin A_{\bar{r}h2}. \end{cases}$$

Similarly the full jackknife variance estimator of (6) is

$$\hat{V}_{J3} = \sum_{k \in A_1} c_k (t_3^{(k)} - t_3)^2, \tag{13}$$

where $t_3^{(k)} = \frac{1}{n^*}\sum_{h=1}^{L} n_h^* \bar{y}_{h2}^{(k)}$
where

$$\frac{1}{n^*}n_h^* = \begin{cases} (n-1)^{-1}(n_h-1) & if \;\; k \in A_{h1} \\ (n-1)^{-1}n_h & if \;\; k \notin A_{h1}, \end{cases}$$

and

$$\bar{y}_{h2}^{(k)} = \begin{cases} (n'_{rh}-1)^{-1}(n'_{rh}\bar{y}_{h2}-y_k) & if \;\; k \in A_{rh2} \\ \bar{y}_{h2} & if \;\; k \notin A_{rh2}. \end{cases}$$

Also the full jackknife variance estimator of (9) is

$$\hat{V}_{J4} = \sum_{k \in A_1} c_k (t_4^{(k)} - t_4)^2, \tag{14}$$

where $t_4^{(k)} = \frac{1}{n^*}\sum_{h=1}^{L} u_{h1}^* n_h^* \bar{y}_{rh2}^{(k)} + \frac{1}{n^*}\sum_{h=1}^{L} u_{h1}^* n_h^* \bar{y}_{\bar{r}h2}^{(k)}$
where

$$\frac{1}{n^*}n_h^* = \begin{cases} (n-1)^{-1}(n_h-1) & if \;\; k \in A_{h1} \\ (n-1)^{-1}n_h & if \;\; k \notin A_{h1}, \end{cases}$$

$$\bar{y}_{rh2}^{(k)} = \begin{cases} (n'_{rh}-1)^{-1}(n'_{rh}\bar{y}_{rh2}-y_{rk}) & if \;\; k \in A_{h2} \\ \bar{y}_{rh2} & if \;\; k \notin A_{h2}, \end{cases}$$

$$\bar{y}_{\bar{r}h2}^{(k)} = \begin{cases} (n'_{\bar{r}h}-1)^{-1}(n'_{\bar{r}h}\bar{y}_{\bar{r}h2}-y_{\bar{r}k}) & if \;\; k \in A_{h2} \\ \bar{y}_{\bar{r}h2} & if \;\; k \notin A_{h2}, \end{cases}$$

and $[u_{\mathrm{h}1}^* = (n-1)^{-1}(n_h-1)]$.

4. Extension of Proposed Estimators

Here, we consider some extensions of the proposed replication method by DEE and REE under the conditions defined in section 2 for proposed estimators.

4.1. DEE and REE in the Presence of Non-response at First Phase Sample

The DEE and REE type estimators were developed for the two situations which was discussed in the section 2.1.

4.1.1. Estimation by Ignoring Non-responding Units

Let $w_{ri} = \pi_{1i}^{-1}$where $\pi_{1i} = \Pr(i \in A_{rh1})$ is the first phase sampling wight for responding units and $w_{ri}^* = \pi_{2i}^{-1}$ where $\pi_{2i} = \Pr(i \in A_{rh2}|i \in A_{rh1})$ be the inverse of the conditional probability of responding units in the second phase, then we develop the following DEE and REE.

The proposed DEE for this situation is

$$t_{1DEE} = \frac{1}{N}\sum_{h=1}^{L}\sum_{i \in A_{rh2}} w_{ri} w_{ri}^* y_i, \tag{15}$$

the replication variance estimator of (15) is developed by the Kim and Yu (2011) method which is as

$$\hat{V}_{J_{1DEE}} = \sum_{k \in A_1} c_k \left(t_{1DEE}^{(k)} - t_{1DEE} \right)^2, \tag{16}$$

where
$t_{1DEE}^{(k)} = \frac{1}{N} \sum_{h=1}^{L} \left(\frac{\sum_{i \in A_{h1}} w_{ri}^{(k)} w_{ri}^{-1}}{\sum_{i \in A_{h2}} w_{ri}^{(k)} w_{ri}^{-1}} \right) \sum_{i \in A_{h2}} w_{ri}^{(k)} y_i$
and c_k is a factor associated with replicate k determined by the replication method. The proposed REE estimator for this situation is

$$t_{1REE} = \frac{1}{N} \sum_{h=1}^{L} \widehat{w}_{h1} \, \bar{y}_{h2}, \tag{17}$$

where
$\bar{y}_{h2} = \left(\sum_{i \in A_{rh2}} w_{ri} w_{ri}^{*} \right)^{-1} \sum_{i \in A_{rh2}} w_{ri} w_{ri}^{*} y_i$ and $\widehat{w}_{h1} = \sum_{i \in A_{rh1}} w_{ri}$.
The replication variance estimator of (17) is as

$$\hat{V}_{J_{1REE}} = \sum_{k \in A_1} c_k \left(t_{1REE}^{(k)} - t_{1REE} \right)^2, \tag{18}$$

where
$t_{1REE}^{(k)} = \frac{1}{N} \sum_{h=1}^{L} \widehat{w}_{h1}^{(k)} \, \bar{y}_{h2}^{(k)}$
and $\widehat{w}_{h1}^{(k)} = \sum_{i \in A_{rh2}} w_{ri}^{(k)}$, $\bar{y}_{h2}^{(k)} = \left(\sum_{i \in A_{rh2}} w_{ri}^{(k)} w_{ri}^{*} \right)^{-1} \sum_{i \in A_{rh2}} w_{ri}^{(k)} w_{ri}^{*} y_i$.

4.1.2. Estimation by Re-contacting Non-responding Units

Let we consider w_{ri} and w_{ri}^{*} as given in Situation-I, further we let $w_{\bar{r}\,i} = \pi_{\bar{r}1i}^{-1}$ where $\pi_{\bar{r}1i} = \Pr(i \in A_{\bar{r}h1})$ is the first phase sampling weight for re-contacted responding units and $w_{\bar{r}\,i}^{*} = \pi_{\bar{r}2i}^{-1}$ where $\pi_{\bar{r}2i} = \Pr(i \in A_{\bar{r}h2} | i \in A_{\bar{r}h1})$ be the inverse of the conditional probability in the second phase of re-contacted responding units, then we develop the following DEE and REE.

The proposed DEE for this situation is

$$t_{2DEE} = \frac{1}{N} \sum_{h=1}^{L} \left(\sum_{i \in A_{rh2}} w_{ri} w_{ri}^{*} y_{ri} + \sum_{i \in A_{\bar{r}h2}} w_{\bar{r}\,i} w_{\bar{r}\,i}^{*} y_{\bar{r}\,i} \right) \tag{19}$$

and replication variance estimator of (19) is as

$$\hat{V}_{J_{2DEE}} = \sum_{k \in A_1} c_k \left(t_{2DEE}^{(k)} - t_{2DEE} \right)^2, \tag{20}$$

where

$$t_{2DEE}^{(k)} = \frac{1}{N} \sum_{h=1}^{L} \left[\left(\frac{\sum_{i \in A_{rh1}} w_{ri}^{(k)} w_{ri}^{-1}}{\sum_{i \in A_{rh2}} w_{ri}^{(k)} w_{ri}^{-1}} \right) \sum_{i \in A_{rh2}} w_{ri}^{(k)} y_{ri} \right.$$

$$+\left(\frac{\sum_{i\in A_{\bar{r}h1}} w_{\bar{r}\,i}^{(k)} w_{\bar{r}\,i}^{-1}}{\sum_{i\in A_{\bar{r}h2}} w_{\bar{r}\,i}^{(k)} w_{\bar{r}\,i}^{-1}}\right)\sum_{i\in A_{\bar{r}h2}} w_{\bar{r}i}^{(k)} y_{\bar{r}\,i}\Bigg]$$

and c_k is a factor associated with replicate k determined by the replication method. The proposed REE for this situation is

$$t_{2REE}=\frac{1}{N}\sum_{h=1}^{L}\left(\widehat{w}_{h1}\ \bar{y}_{rh2}+\widehat{w}^{*}_{h1}\ \bar{y}_{\bar{r}h2}\right), \tag{21}$$

where
$\widehat{w}_{h1}=\sum_{i\in A_{rh1}} w_{ri}$, $\bar{y}_{rh2}=\left(\sum_{i\in A_{rh2}} w_{ri}w^{*}_{ri}\right)^{-1}\sum_{i\in A_{rh2}} w_{ri}w^{*}_{ri}y_{ri}$,
$\widehat{w}^{*}_{h1}=\sum_{i\in A_{rh1}} w_{\bar{r}i}$ and $\bar{y}_{\bar{r}h2}=\left(\sum_{i\in A_{\bar{r}h2}} w_{\bar{r}i}w^{*}_{\bar{r}i}\right)^{-1}\sum_{i\in A_{\bar{r}h2}} w_{\bar{r}i}w^{*}_{\bar{r}i}y_{\bar{r}i}$.
The replication variance estimator of (21) is

$$\hat{V}_{J_{2REE}}=\sum_{k\in A_1} c_k\left(t^{(k)}_{2REE}-t_{2REE}\right)^2, \tag{22}$$

where
$t^{(k)}_{2REE}=\frac{1}{N}\sum_{h=1}^{L}\left(\widehat{w}^{(k)}_{h1}\ \bar{y}^{(k)}_{rh2}+\widehat{w}^{(k)}_{h1}\ \bar{y}^{(k)}_{\bar{r}h2}\right)$
where
$\bar{y}^{(k)}_{rh2}=\left(\sum_{i\in A_{rh2}} w^{(k)}_{ri}w^{*}_{ri}\right)^{-1}\sum_{i\in A_{rh2}} w^{(k)}_{ri}w^{*}_{ri}y_{ri}$, $\widehat{w}^{(k)}_{h1}=\sum_{i\in A_{rh2}} w^{(k)}_{ri}$,
$\widehat{w}^{*(k)}_{h1}=\sum_{i\in A_{\bar{r}h2}} w^{(k)}_{\bar{r}i}$ and $\bar{y}^{(k)}_{\bar{r}h2}=\left(\sum_{i\in A_{\bar{r}h2}} w^{(k)}_{\bar{r}i}w^{*}_{\bar{r}i}\right)^{-1}\sum_{i\in A_{\bar{r}h2}} w^{(k)}_{\bar{r}i}w^{*}_{\bar{r}i}y_{\bar{r}i}$.

4.2. DEE and REE in the presence of non-response at second phase

In this section we again develop two new estimators of DEE and REE for the two situations discussed in section 2.2.

4.2.1. Estimation by ignoring non-responding units

Let $w_i=\pi_i^{-1}$ where $\pi_i=\Pr(i\in A_{h1})$ is the first phase sampling weight and $w^{*}_i=\pi_{ri}^{-1}$ where $\pi_{ri}=\Pr(i\in A_{rh2}|i\in A_{h1})$, be the inverse of the conditional probability in the second phase responding units, then we developed DEE and REE by ignoring second phase non-responding units.

The proposed DEE estimator for this situation is

$$t_{3DEE}=\frac{1}{N}\sum_{h=1}^{L}\sum_{i\in A_{h2}} w_i w^{*}_i y_i \tag{23}$$

and replication variance estimator of (23) is

$$\hat{V}_{J_{3DEE}}=\sum_{k\in A_1} c_k\left(t^{(k)}_{3DEE}-t_{3DEE}\right)^2, \tag{24}$$

where

$t_{3DEE}^{(k)}=\frac{1}{N}\sum_{h=1}^{L}\left(\frac{\sum_{i\in A_{h1}}w_i^{(k)}w_i^{-1}}{\sum_{i\in A_{rh2}}w_i^{(k)}w_i^{-1}}\right)\sum_{i\in A_{rh2}}w_i^{(k)}y_i$

and c_k is a factor associated with replicate k determined by the replication method. The proposed REE estimator for this situation is

$$t_{3REE}=\frac{1}{N}\sum_{h=1}^{L}\widehat{w}_{h1}\;\bar{y}_{rh2}, \tag{25}$$

where

$\widehat{w}_{h1}=\sum_{i\in A_{h1}}w_i$ and $\bar{y}_{h2}=\left(\sum_{i\in A_{rh2}}w_iw_i^*\right)^{-1}\sum_{i\in A_{rh2}}w_iw_i^*y_{ri}$.

The replication variance estimator of (25) is

$$\hat{V}_{J_{3REE}}=\sum_{k\in A_1}c_k\left(t_{3REE}^{(k)}-t_{3REE}\right)^2, \tag{26}$$

where

$t_{3REE}^{(k)}=\frac{1}{N}\sum_{h=1}^{L}\widehat{w}_{h1}^{(k)}\;\bar{y}_{rh2}^{(k)}$,

and

$\bar{y}_{rh2}^{(k)}=\left(\sum_{i\in A_{rh2}}w_i^{(k)}w_i^*\right)^{-1}\sum_{i\in A_{rh2}}w_i^{(k)}w_i^*y_{ri}\;,\;\widehat{w}_{h1}^{(k)}=\sum_{i\in A_{h1}}w_i^{(k)}$.

4.2.2. Estimation by re-contacting non responding units

Consider w_i and w_i^* as given in Situation-I, further we let $w_{\bar{r}^*i}=\pi_{\bar{r}^*i}^{-1}$ where $\pi_{\bar{r}^*i}=\Pr(i\in A_{\bar{r}^*h2}|i\in A_{h1})$ be the inverse of the conditional probability in the second phase re-contacted units, then by re-contacting the non-responding units we develop DEE and REE.

The proposed DEE estimator for this situation is

$$t_{4DEE}=\frac{1}{N}\sum_{h=1}^{L}\left(\sum_{i\in A_{h2}}w_iw_i^*y_{ri}+\sum_{i\in A_{h2}}w_iw_{\bar{r}^*i}y_{\bar{r}^*i}\right) \tag{27}$$

and replication variance estimator of (27) is

$$\hat{V}_{J_{4DEE}}=\sum_{k\in A_1}c_k\left(t_{4DEE}^{(k)}-t_{4DEE}\right)^2, \tag{28}$$

where

$$t_{4DEE}^{(k)}=\frac{1}{N}\sum_{h=1}^{L}\Bigg[\left(\frac{\sum_{i\in A_{h1}}w_i^{(k)}w_i^{-1}}{\sum_{i\in A_{h2}}w_i^{(k)}w_i^{-1}}\right)\sum_{i\in A_{h2}}w_i^{(k)}y_{ri} + \left(\frac{\sum_{i\in A_{h1}}w_i^{(k)}w_i^{-1}}{\sum_{i\in A_{h2}}w_i^{(k)}w_i^{-1}}\right)\sum_{i\in A_{h2}}w_i^{(k)}y_{\bar{r}^*i}\Bigg]$$

and c_k is a factor associated with replicate k determined by the replication method.

The proposed REE estimator for this situation is

$$t_{4REE} = \frac{1}{N}\sum_{h=1}^{L} \widehat{w}_{h1}\left(\bar{y}_{rh2} + \bar{y}_{\bar{r}h2}\right), \tag{29}$$

where
$\widehat{w}_{h1} = \sum_{i\in A_{h1}} w_i$, $\bar{y}_{rh2} = \left(\sum_{i\in A_{h2}} w_i w_i^*\right)^{-1} \sum_{i\in A_{h2}} w_i w_i^* y_{ri}$,
$\bar{y}_{\bar{r}^*h2} = \left(\sum_{i\in A_{h2}} w_i w_{\bar{r}^*i}\right)^{-1} \sum_{i\in A_{h2}} w_i w_{\bar{r}^*i} y_{\bar{r}^*i}$.

The replication variance estimator of (29) is

$$\hat{V}_{J_{4REE}} = \sum_{k\in A_1} c_k \left(t_{4REE}^{(k)} - t_{4REE}\right)^2, \tag{30}$$

where
$t_{4REE}^{(k)} = \frac{1}{N}\sum_{h=1}^{L} \widehat{w}_{h1}^{(k)} \left(\bar{y}_{rh2}^{(k)} + \bar{y}_{\bar{r}^*h2}^{(k)}\right)$
and $\bar{y}_{\bar{r}^*h2}^{(k)} = \left(\sum_{i\in A_{h2}} w_i^{(k)} w_{\bar{r}^*i}\right)^{-1} \sum_{i\in A_{h2}} w_i^{(k)} w_{\bar{r}^*i} y_{\bar{r}^*i}$, $\widehat{w}_{h1}^{(k)} = \sum_{i\in A_{h1}} w_i^{(k)}$,
$\bar{y}_{rh2}^{(k)} = \left(\sum_{i\in A_{h2}} w_i^{(k)} w_i^*\right)^{-1} \sum_{i\in A_{h2}} w_i^{(k)} w_i^* y_{ri}$.

5. Simulation Study

We conducted a simulation study, to analyze the performance of the estimators proposed in Section 2 to Section 4. A finite artificial population of size N = 1000 was generated with four variables (y_i, z_i, q_i, u_i), where the population elements are independently generated from $z_i \sim \exp(1)+2$; $q_i \sim \chi^2(1)+2$; $u_i \sim Unif\,(1, 2, 3)$ and $y_i = \beta_0 + \beta_1 z_i + \beta_2 q_i + \varepsilon_i$ where β_0, β_1, β_2 are 0, 1, 1 respectively and $\varepsilon_i \sim N(0, 1)$ as discussed by Kim and Yu (2011). To obtain unequal probability samples for this simulation study, we used SRSWOR, with selection probabilities proportional to the measure of size variable. The variable z_i was used as a size measure for the unequal probability sampling in the first phase sampling and q_i was used as a size measure for the unequal probability sampling in the second phase sampling, the stratification was defined using variable u_i, it is assumed that the variable z_i, q_i, u_i, y_i and ε_i are mutually independent.

A first phase sample of size $n = 200$ was drawn from the population of 1000. For first situation of section 2.1 it is assumed that 30% non-response at first phase. First phase responding sample of size n_r was stratified using u_i and stratified random sample of size n' was taken at second phase. In situation-II first phase sampled n units are stratified using u_i for second phase sampling for both assumed responding and re-contacting elements of first phase sample and stratified random sample of size $n' = n'_r + n'_{\bar{r}}$ was taken.

When non-response is present on second phase, then for situation-I discussed in section 2.2 a sample of size $n = 200$ with SRSWOR was drawn. First phase sampled n units are stratified into L stratum. It is assumed that 30% non-responses exist in second phase sample, so ignoring non-responding units a second phase stratified sample of size n'_{rh} was taken from h^{th}stratum. For situation-II,the estimator is developed by re-contacting second phase non-responding units and a stratified random sample of size $n' = \sum_{h=1}^{L} n'_{rh} + n'_{\bar{r}h}$ is used for both assumed responding and re-contacting elements of second phase sample.

We generate S=4000 independent samples for simulation from the finite population generated above. the parameter of interest is the population mean of the variable y. From generated samples, we compute two phase stratified estimators, DEE and REE with their replication variances. Relative bias (RB) and coefficient of variation (CV) were computed for each estimator and their replication variance with the following method.

1. The percent relative bias of the mean estimators with respect to the population parameter

$$PRB = \frac{E(t) - T}{T} \times 100, \tag{31}$$

2. The percent relative bias of the jackknife variance estimators with respect to the estimated true mean squared error is estimated by

$$PRB = \frac{E(V_J(t)) - MSE_{true}}{MSE_{true}} \times 100, \tag{32}$$

where $E(V_J(t)) = \frac{1}{S}\sum_{i=1}^{S} V_J(t_i)$ and $MSE_{true} = \frac{1}{S}\sum_{i=1}^{S}(t_i - T)^2$.

3. The percent coefficient of variation of the jackknife variance with respect to the estimated true MSE is estimated by

$$PCV = \frac{\sqrt{\frac{1}{S}\sum_{i=1}^{S}(V_{Ji}(t) - MSE_{true})^2}}{MSE_{true}} \times 100. \tag{33}$$

Results of proposed estimators with percent relative bias and percent coefficient of variation are given in Table 1. Table 2 shows the percent relative bias and percent coefficient of variation of the replication variance estimators.

6. Concluding Remarks

Table 1 shows that estimators are very close values to population parameter with small relative bias. Table 2 shows the RB and CV of the replication variance of the proposed estimators. The result shows that the variances are with smaller bias. The main purpose of this study was to show that the replication variance estimator works well in two phase stratified design when non-response is present. Expansion estimators are used in the estimation strategy and simulation study supported these results.

ACKNOWLEDGEMENTS

The authors are highly thankful to Dr Giovanna Galatà– Executive Editor Support and to the referee for their valuable suggestions regarding improvement of the paper.

REFERENCES

B. B. Khare, S. Srivastava (1997). *Transformed ratio type estimators for population mean in presence of non-response*, Commu. Statist. Theory Meth., 26, pp. 1779-1791.

H. L. Jones (1958). *The jackknife method*, Proceedings of the IBM scientific computation and symposium on statistics, pp. 185-197.

H. P. Singh, L. N. Upadhyaya, R. Iachan (1990). *On applying the jackknife technique to the ratio and product estimators in two-phase sampling*, Vikram Math. J., X, pp. 45-58.

J. Breidt, W.A. Fuller (1993).*Regression weighting for multiplicative samples*, Sankhya, 55, pp. 297-309.

J. Durbin (1959). *A note on the application of Quenouille's method of bias reduction to the estimation of ratios*, Biometrika, 46, pp. 477-480.

J. K. Kim, A. Navarro, W. A. Fuller (2006). *Replication variance estimation for two phase stratified sampling*, Journal of the American statistical association, 101, pp. 312-320.

J. K. Kim, R. R. Sitter (2003). *Efficient replication variance estimation for two phase sampling*, Statistica Sinica, 13, pp. 641-653.

J. K. Kim, C. L. Yu (2011). *Replication variance estimation under two-phase sampling*, Survey methodology, 37, pp. 67-74.

J. Neyman (1938). *Contribution to the theory of sampling human populations*, Journal of the American Statistical Association, 33, pp. 101-116.

J. N. K. Rao (1973). *On double sampling for stratification and analytical surveys*, Biometrika, 60, pp. 125-133.

J. N. K. Rao, J. Shao (1992). *Jackknife variance estimation with survey data under hot deck imputation*, Biometrika, 79, pp. 811-822.

J. N. K. Rao, R. R. Sitter (1995). *Variance estimation under two-phase sampling with application to imputation for missing data*, Biometrika, 82, pp. 453-460.

J. Tukey (1958). *Bias and Confidence in not quite large samples*, Ann. Math. Statist., 29, pp. 614.

K. F. Rust, J. N. K. Rao (1995). *Variance estimation under two phase sampling with application to imputation for missing data*, Biometrika, 82, pp. 453-460.

K. Wolter (2007). *Introduction to Variance Estimation. 2^{nd} Edition*, New York: Springer.

L. Chand (1975). *Some ratio type estimators based on two or more auxiliary variables*, PhD Thesis, Iowa State Unversity, Ames, Iowa, USA.

M. H. Hansen, W. N. Hurwits (1946). *The problem of non-response in sample surveys*, Journal of the American Statistical Association, 41, pp. 517-529.

M. H. Quenouille (1956). *Notes on bias in estimation*, Biometrika, 43, pp. 353-360.

P. J. Farrell, S. Singh (2010). *Some Contributions to Jackknifing Two-phase Sampling Estimators*, Survey Methodology, 36, pp. 57-68.

P. S. Kott (1990). *Variance estimation when a first-phase area sample is restratified*, Survey Methodology, 16, pp. 99-103.

P. S. Kott, D. M. Stukel (1997). *Can the jackknife be used with a two phase sample*, Survey Methodology, 23, pp. 81-89.

P. S. R. S. Rao (1986). *Ratio estimation with subsampling the non-respondents*, Survey Meth., 12, pp. 217-230.

R. R. Sitter (1997). *Variance estimation for the regression estimator in two-phase sampling*, Journal of the American Statistical Association, 92, pp. 780-787.

V. Ramasubramanian, A. Rai, R. Sing (2007). *Jackknife variance estimation under two phase sampling*, Model Assisted Statistics and Application, 2, pp. 27-35.

W. G. Cochran (1977). *Sampling Techniques*, New York: John Wiley & Sons, lnc.

W. A. Fuller (1998). *Replication variance estimation for two phase samples*, Sta- tistica sinica, 8, pp. 1153-1164.

SUMMARY

Replication Variance Estimation under Two-phase Sampling in the Presence of Non-response

Kim and Yu (2011) discussed replication variance estimator for two-phase stratified sampling. In this paper estimators for mean have been proposed in two-phase stratified sampling for different situation of existence of non-response at first phase and second phase. The expressions of variances of these estimators have been derived. Furthermore, replication-based jackknife variance estimators of these variances have also been derived. Simulation study has been conducted to investigate the performance of the suggested estimators.

Keywords: Two-phase stratified sampling; Jackknife variance estimator; Non-response

Appendix

Proof of (2): As we know in the case of two phase sampling, in general the variance of an estimator can be expressed as

$$V(t_1) = V_1\left[E_2(t_1)\right] + E_1\left[V_2(t_1)\right]. \tag{34}$$

Considering $E_1\left[V_2(t_1)\right]$ from above Eq. (34)

$$\begin{aligned} E_1\left[V_2\left(t_1\right)\right] =&\ E_1\left[V_2\left(\frac{1}{n_r}\sum_{h=1}^{L}\sum_{i\in A_{rh2}}\frac{n_{rh}}{n'_{rh}}y_{rhi}\right)\right] \\ =&\ E_1\left[V_2\left(\frac{1}{n_r}\sum_{h=1}^{L}n_{rh}\bar{y}_{rh}\right)\right] \end{aligned}$$

or

$$E_1\left[V_2(t_1)\right] = E\left[\sum_{h=1}^{L}\left(\frac{n_{rh}}{n_r}\right)^2\left(\frac{1}{n'_{rh}} - \frac{1}{n_{rh}}\right)s^2_{rh1}\right], \tag{35}$$

where $s^2_{rh1} = \left(n_{rh} - 1\right)^{-1}\sum\limits_{i\in A_{rh1}}\left(y_i - \bar{y}_{rh1}\right)^2$ and $\bar{y}_{rh1} = n_{rh}^{-1}\sum\limits_{i\in A_{rh1}}y_{ri}$.
Now considering $V_1\left[E_2(t_1)\right] from Eq.$ (34)

$$\begin{aligned} V_1\left[E_2(t_1)\right] &= V_1\left[E_2\left(\frac{1}{n_r}\sum_{h=1}^{L}n_{rh}\bar{y}_{rh}\right)\right] \\ &= V_1\left(\bar{y}\right) \\ &= \frac{N - n_r}{n_r N}S^2 \end{aligned}$$

or

$$V_1\left[E_2(t_1)\right] = n_r^{-1}\left(1 - f_1\right)S^2 \tag{36}$$

where $f_1 = \frac{n_r}{N}, S^2 = \left(N-1\right)^{-1}\sum\limits_{i=1}^{N}\left(y_i - \bar{y}_N\right)^2$.
Using Eq. (35) and Eq. (36) in Eq. (34), we get

$$V\left(t_1\right) = n_r^{-1}\left(1 - f_1\right)S^2 + E\left[\sum_{h=1}^{L}\left(\frac{n_{rh}}{n_r}\right)^2\left(\frac{1}{n'_{rh}} - \frac{1}{n_{rh}}\right)s^2_{rh1}\right].$$

As

$$n_r^{-1}S^2 = E\left[n_r^{-1}\sum_{h=1}^{L}w_h\left(\left(\bar{y}_{rh1} - \bar{y}_1\right)^2 + s^2_{rh1}\right)\right].$$

Then

$$V(t_1) = \left(1 - f_1\right)E\left[n_r^{-1}\sum_{h=1}^{L}w_h\left(\left(\bar{y}_{rh1} - \bar{y}_1\right)^2 + s^2_{h1}\right)\right]$$

$$+E\left[\sum_{h=1}^{L}\left(\frac{n_{rh}}{n_r}\right)^2\left(\frac{1}{n'_{rh}}-\frac{1}{n_{rh}}\right)s^2_{rh1}\right]$$

$$=E\left[n_r^{-1}\left(1-f_1\right)\sum_{h=1}^{L}w_h\left(\bar{y}_{rh1}-\bar{y}_1\right)^2+n_r^{-1}\sum_{h=1}^{L}w_h s^2_{rh1}\right.$$

$$\left.-n_r^{-1}f_1\sum_{h=1}^{L}w_h s^2_{rh1}+\sum_{h=1}^{L}{n'}^{-1}_{rh}w_h^2 s^2_{rh1}-\sum_{h=1}^{L}n^{-1}_{rh}w_h^2 s^2_{rh1}\right]$$

or

$$V(t_1)=E\left[n_r^{-1}\left(1-f_1\right)\sum_{h=1}^{L}w_h\left(\bar{y}_{rh1}-\bar{y}_1\right)^2-n_r^{-1}f_1\sum_{h=1}^{L}w_h s^2_{rh1}\right.$$

$$\left.+\sum_{h=1}^{L}{n'}^{-1}_{rh}w_h^2 s^2_{rh1}\right].$$

Solving above equation we can obtain required result given in Eq.(2).

Proof of (5): Estimator given in (4) can be written as

$$t_2=w_1t_1^*+w_2t_2^*, \tag{37}$$

where $t_1^*=\frac{1}{n_r}\sum_{h=1}^{L}\sum_{i\in A_{rh2}}\frac{n_{rh}}{n'_{1h}}y_{rhi}$, $t_2^*=\frac{1}{n_{\bar{r}}}\sum_{h=1}^{L}\sum_{i\in A_{\bar{r}h2}}\frac{n_{\bar{r}h}}{n'_{2h}}y'_{rhi}$, $w_1=\frac{n'_1}{n_r}$ and $w_2=\frac{n'_2}{n_{\bar{r}}}$.

As t_1^* and t_2^* are independent, the variance expression of Eq. (37) can be written as

$$V\left(t_2\right)=w_1^2V\left(t_1^*\right)+w_2^2V\left(t_2^*\right). \tag{38}$$

Considering $V\left(t_1^*\right)$ from Eq. (38), we can write

$$V(t_1^*)=V_1\left[E_2(t_1^*)\right]+E_1\left[V_2(t_1^*)\right]. \tag{39}$$

Now considering $[E_1\left[V_2\left(t_1^*\right)\right]$ from Eq. (39), we can write

$$E_1\left[V_2\left(t_1^*\right)\right]=E_1\left[V_2\left(\frac{1}{n_r}\sum_{h=1}^{L}n_{rh}\bar{y}_{rh}\right)\right]$$

or

$$E_1\left[V_2\left(t_1^*\right)\right]=E\left[\sum_{h=1}^{L}\left(\frac{n_{rh}}{n_r}\right)^2\left(\frac{1}{n'_{1h}}-\frac{1}{n_{rh}}\right)s^2_{rh1}\right], \tag{40}$$

where $s^2_{rh1}=\left(n_{rh}-1\right)^{-1}\sum_{i\in A_{rh1}}\left(y_{ri}-\bar{y}_{rh1}\right)^2$ and$\bar{y}_{rh1}=n^{-1}_{rh}\sum_{i\in A_{rh1}}y_{ri}$.
Now considering $V_1\left[E_2(t_1^*)\right]$ from Eq. (39), we can write

$$V_1\left[E_2\left(t_1^*\right)\right]=V_1\left[E_2\left(\frac{1}{n_r}\sum_{h=1}^{L}n_{rh}\bar{y}_{rh}\right)\right]$$

$$= V_1\left(\bar{y}\right)$$

or

$$V_1\left[E_2\left(t_1^*\right)\right] = \; n_r^{-1}\left(1-f_1\right)S^2 \tag{41}$$

where $f_1 = \frac{n_r}{N} S^2 = (N-1)^{-1}\sum_{i=1}^{N}\left(y_i - \bar{y}_N\right)^2$.
Using Eq. (40) and Eq. (41) in Eq. (39), we get

$$\hat{V}\left(t_1^*\right) = \; n_r^{-1}\left(1-f_1\right)\sum_{h=1}^{L} w_h\left(\bar{y}_{rh2}-\bar{y}_{r2}\right)^2 + \sum_{h=1}^{L}\left({n'}_{1h}^{-1} - n_{rh}^{-1}f_1\right) w_h^2 s_{rh2}^2. \tag{42}$$

From Eq. (38), now we consider

$$V\left(t_2^*\right) = V_1\left[E_2\left(t_2^*\right)\right] + E_1\left[V_2\left(t_2^*\right)\right] \tag{43}$$

Now considering $E_1\left[V_2\left(t_2^*\right)\right]$ from Eq. (43), we can write

$$E_1\left[V_2\left(t_2^*\right)\right] = \; E_1\left[V_2\left(\frac{1}{n_{\bar{r}}}\sum_{h=1}^{L} n_{\bar{r}h}\bar{y}_{\bar{r}h}\right)\right]$$

or

$$E_1\left[V_2(t_2^*)\right] = \; E\left[\sum_{h=1}^{L}\left(\frac{n_{\bar{r}h}}{n_{\bar{r}}}\right)^2\left(\frac{1}{n'_{2h}} - \frac{1}{n_{\bar{r}h}}\right)s_{\bar{r}h1}^2\right] \tag{44}$$

Now considering $V_1\left[E_2\left(t_2^*\right)\right]$ from Eq. (43), we can write

$$V_1\left[E_2\left(t_2^*\right)\right] = \; V_1\left[E_2\left(\frac{1}{n_{\bar{r}}}\sum_{h=1}^{L} n_{\bar{r}h}\bar{y}_{\bar{r}h}\right)\right]$$

$$= \; V_1\left[\bar{y}\right]$$

or

$$V_1\left[E_2\left(t_2^*\right)\right] = \; n_{\bar{r}}^{-1}\left(1-f_2\right)S^2 \tag{45}$$

Using Eq. (44) and Eq. (45) in Eq. (43), we get

$$\hat{V}\left(t_2^*\right) = \; n_{\bar{r}}^{-1}\left(1-f_2\right)\sum_{h=1}^{L} w_h^*\left(\bar{y}_{\bar{r}h2}-\bar{y}_{\bar{r}2}\right)^2 + \sum_{h=1}^{L}\left({n'}_{2h}^{-1} - n_{\bar{r}h}^{-1}f_2\right) w_h^{*2} s_{\bar{r}h2}^2 \tag{46}$$

Using Eq. (42) and Eq. (46) in Eq. (37), we get

$$\hat{V}\left(t_2\right) = \; w_1^2\left(n_r^{-1}\left(1-f_1\right)\sum_{h=1}^{L} w_h\left(\bar{y}_{rh2}-\bar{y}_{r2}\right)^2 + \sum_{h=1}^{L}\left({n'}_{1h}^{-1} - n_{rh}^{-1}f_1\right) w_h^2 s_{rh2}^2\right)$$

$$+\, w_2^2\left(n_{\bar{r}}^{-1}\left(1-f_2\right)\sum_{h=1}^{L} w_h^*\left(\bar{y}_{\bar{r}h2}-\bar{y}_{\bar{r}2}\right)^2 + \sum_{h=1}^{L}\left({n'}_{2h}^{-1} - n_{\bar{r}h}^{-1}f_2\right) w_h^{*2} s_{\bar{r}h2}^2\right) \tag{47}$$

Solving Eq. (47) we can obtain required result given in Eq.(5).

TABLE 1
RB and CV of Proposed Estimators

Estimator	Parameter	Estimated Mean	RB	CV
t_1	2.32	2.35	3.95	7.25
t_2	2.32	2. 43	11.73	11.52
t_3	2.32	2.55	- 5.21	10.87
t_4	2.32	2.60	- 7.85	16.99
t_{1DEE}	2.32	2.50	7.27	20.74
t_{2DEE}	2.32	2.39	19.27	31.07
t_{3DEE}	2.32	2.21	14.12	29.42
t_{4DEE}	2.32	2.43	18.73	17.13
t_{1REE}	2.32	2.55	10.38	39.72
t_{2REE}	2.32	2.48	-5.73	27.11
t_{3REE}	2.32	2.38	6.11	36.21
t_{4REE}	2.32	2.47	15.13	42.38

TABLE 2
RB and CV of Replication Variance Estimators

Estimator	RB	CV
$V_J\,(t_1)$	19.22	31.83
$V_J\,(t_2)$	17.03	28.53
$V_J\,(t_3)$	15.66	18.13
$V_J\,(t_4)$	25.55	53.29
$V_J\,(t_{1DEE})$	11.12	17.52
$V_J\,(t_{2DEE})$	26.48	59.71
$V_J\,(t_{3DEE})$	22.14	36.87
$V_J\,(t_{4DEE})$	38.26	41.30
$V_J\,(t_{1REE})$	11.75	17.22
$V_J\,(t_{2REE})$	13.63	22.87
$V_J\,(t_{3REE})$	9.28	15.81
$V_J\,(t_{4REE})$	12.20	29.38