Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online): 2229 5666 Vol. 11, Issue 1 (2018): 51-66

ESTIMATION OF FINITE POPULATION MEAN USING KNOWN COEFFICIENT OF VARIATION IN THE SIMULTANEOUS PRESENCE OF NON - RESPONSE AND MEASUREMENT ERRORS UNDER DOUBLE SAMPLING SCHEME

[1]Dharmendra Kumar Yadav, Monika Devi[*] and [3]Subhash Kumar Yadav
[1]Deptt. of Statistics, Ramanujan College, University of Delhi, New Delhi, India
[*]Sri Karan Narendra Agriculture University, Jobner, Jaipur, India
[3]Deptt. of Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow, India
E Mail: [1]dkumar.yadava@gmail.com; [*]mscagstats@gmail.com;
[3]drskystats@gmail.com
[*]Corresponding Author

Received February 19, 2018
Modified May 11, 2018
Accepted June 11, 2018

Abstract

The present article considers the problem of finite population mean estimation, when the non-response and measurement errors are present simultaneously utilizing known information on coefficient of variation of study variable. We have developed an estimator of population mean which is improved and efficient, using Hansen and Hurwitz (1946) technique. Asymptotic expressions of the bias and variance of suggested estimator have been found correct up to approximation of degree one.The optimum value of characterizing scalar for which variance of proposed estimator is minimum has also been calculated. We have made a theoretical efficiency comparison of proposed estimator with usual Hansen Hurwitz estimator. To amply corroborate the theoretical findings, a simulation study has also been carried out using R software.

Key Words: Estimation, Coefficient of Variation, Non- response, Measurement Errors, Double Sampling, Simulation.

1. Introduction

In sampling theory it is a very common assumption that the observations taken over the units under consideration are measured correctly on characteristics under study. But in practice, it has been seen that this assumption is hardly satisfied in various real world applications and data get contaminated with many errors namely frame errors, non- response, and measurement errors. The errors of measurement also referred to as response errors usually occur at the time of data collection due to discrepancy between observed value and true value of the characteristic under study. These measurement errors make the result invalid and the estimates thus obtained are not reliable. If measurement shows very small variations in errors of measurement so that it can be neglected, the statistical inferences drawn, based on observed data will still be valid. On the other hand, when the measurement errors are not so small that can be neglected, the inferences drawn on the parameters under consideration may be some how accurate and valid but they may often approaches to unexpected, undesirable and unfortunate consequences [Srivastava and Shalabh (2001)].

Cochran (1968), Shalabh (1997), Sud and Srivastva (2000) and Singh and Karpe (2008, 2009, 2010) are some main authors who discussed about various measurement errors. Misra and Yadav (2015) proposed regression type estimator using known coefficient of variation under measurement errors. Misra et al (2016, 2017) have discussed the impact of committing errors in measurement for estimating population mean and population variance in presence of measurement errors in the data obtained from survey sampling.

Consider a finite population $U = (U_1, U_2, \ldots, U_N)$ of N units. Let Y be the study variable. Suppose that we have selected a sample of *n* units by using the technique of simple random sampling without replacement (SRSWOR) on the study variable Y. Further it is assumed that y_i, i=1, 2,…, n is the observed sampling unit having measurement error in place of the true value Y_i of the study variable Y. Let

$$u_i = y_i - Y_i$$

Where, u_i is the measurement error associated with the ith unit of the observation which is random in nature with zero mean and fixed variance σ_u^2 .

Non-response results when respondents do not provide the desired information to the interviewers. Non response rate is represented by the proportion of non-respondents in the sample. It is worth notable that one should take into consideration the maximum response rate as non-respondents may be of different characteristics to respondents. Non-response may lead to the biased and inefficient results of the survey. In surveys sampling, non-response is a very common phenomenon and it is more likely in mail surveys rather than the personal interviews.

Much research work is devoted to consider the problem of non-response in survey sampling. Foradori (1961) studied the sub-sampling of the non-respondents technique for estimating the population total of the study variable in two stages using unequal probability sampling. Tripathi and Khare (1997) generalized the approach of the problem of the sub sampling of non-respondents up to the case of extend of multivariate. Okafor (2001, 2005) further lead this method in element sampling and two-stage sampling respectively for different two successive occasions. Singh and Kumar (2011), Kumar (2014), Singh and Naqvi (2015) and Khare et al (2015) are some remarkable developments in the context of non-response. Devi et al. (2016) suggested the estimation of population mean for two stage sampling under the mechanism of random response in case of non-response.

Hansen and Hurwitz (1946) first took into account the case of non-response in mail surveys, which are commonly used for data collection in advanced countries due to their low cost. The Hansen and Hurwitz (1946) technique consists in selecting sub-sample of initial non-respondents of subsequent data collection with a more expensive method. The technique is, generally, applicable to mail surveys. The problem of non-response is very often in mail surveys. The approach consists in taking a random sub-sample of the persons wh o have not been reached and make a major effort to interview everyone in the sub-sample. It was shown that unbiased estimation is possible despite the non-observation of certain units in the initial sample. The summery of the proposed technique is given below:

1. The questionnaire is mailed to all the selected respondents into the sample.
2. When the time of response is over, the non-respondents are identified and sub-sample of the non- respondents is selected.
3. Data is collected from the non-respondents found in the subsample through conducting the interview.
4. The data so obtained from two parts of the survey is combined for estimating the population parameters.

It is important to mention that the approach used by Hansen and Hurwitz (1946) was based on a deterministic response mechanism. Thus, the population under consideration was supposed to be divided in two different parts, i.e. responding part and non-responding part. Units belonging to the responding part respond with probability one while those in the other part respond with probability zero. Let the population under consideration of size N is divided in two different classes, i.e. the persons responding at the first attempt belong to the response class and the persons not responding at first attempt will be termed as the non-response classes having subpopulation sizes N_1 and N_2 respectively, such that $N_1+N_2=N$. Let y_i be the value of the *i-th* response variable. The population mean $\bar{Y}$ of the study variable can be expressed as,

$$\bar{Y} = \frac{N_1\bar{Y}_1 + N_2\bar{Y}_2}{N} = W_1\bar{Y}_1 + W_2\bar{Y}_2$$

Where W_1 and W_2 are the proportions of units in the response and non-response classes such that $W_1 + W_2 = 1$ and $\bar{Y}_1$ and $\bar{Y}_2$ are the population means in these classes. Thus

$$\bar{Y}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} Y_i \quad \text{and} \quad \bar{Y}_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} Y_i$$

Let n be the size of the simple random sample drawn from the population having N units by simple random sampling without replacement (SRSWOR) technique. Let n_1 out of n denotes the number of units responding in the sample and n_2 units denote the non-responding units. Let h_2 be the size of subsample from the n_2 non-respondents to be interviewed so that $n_2 = h_2 f$. The unbiased estimators of N_1 and N_2 are given by,

Let $\hat{N}_1 = \frac{n_1 N}{n}$, $\hat{N}_2 = \frac{n_2 N}{n}$

Let $\bar{y}_{h_2}$ be the mean of h_2 observations in the sub-sample and define,

$$\bar{y}_w = \frac{n_1\bar{y}_{n_1} + n_2\bar{y}_{h_2}}{n} \tag{1.1}$$

where $\bar{y}_{h_2} = \frac{1}{h_2}\sum_{i=1}^{h_2} y_i$ and $\bar{y}_{n_1} = \frac{1}{n_1}\sum_{i=1}^{n_1} y_i$

The estimator $\bar{y}_w$ is an unbiased estimator of $\bar{Y}$ with variance

$$V(\bar{y}_w) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y{}^2 + \frac{f-1}{n}\frac{N_2}{N} S_{y_2}^2$$

where, $S_{y_2}^2$= population mean square for the non-response class.

For improved estimation of the population mean of the study variable, the important concepts namely measurement error and non-response are merged together. In the present work we have considered that sub sampled part h_2 of non-responding class n_2 contains measurement errors i.e. $u_{ih_2} = y_{ih_2} - Y_{ih_2}$ where u_{ih2} are measurement errors associated with study variable in subsample of non-response class and y_{ih_2} and Y_{ih_2} are observed values and true values respectively.

2. Proposed Estimator for Estimating Population Mean

Misra et al (2017) suggested the given below regression type estimator of population mean of the study variable in presence of measurement errors as,

$$\bar{y}_k = \bar{y} + b(\mu_X - \bar{x}) + k\left(\bar{y} - \frac{\hat{\sigma}_Y}{C_Y}\right) \tag{2.1}$$

where k is the characterising scalar to be chosen suitably

and, $b = \dfrac{\hat{\sigma}_{xy}}{\hat{\sigma}_X^2}$ = Regressioncoefficient

$$\text{Bias}(\bar{y}_k) = \frac{1}{n}\left[\frac{k\mu_Y A_Y}{8} - \frac{\sigma_{XY}}{\sigma_X^2}\left\{\frac{\mu_{2100}}{\sigma_{XY}} - \frac{\mu_{3000}}{\sigma_X^2}\right\}\right]$$

$$\text{MSE}(\bar{y}_k)_{\min} = \frac{\sigma_Y^2}{n}(1-\rho^2) + \frac{1}{n}\left[\sigma_u^2 + \rho^2\sigma_Y^2\left(\frac{\sigma_V^2}{\sigma_X^2}\right)\right] - \frac{\left[\frac{\rho\mu_{1200}}{C_Y\sigma_Y} + 2(\sigma_u^2 + \sigma_Y^2) - \frac{\mu_{0300}}{\sigma_Y C_Y} - 2\rho^2\sigma_X\sigma_Y\right]^2}{n\mu_Y^2\left[\frac{4C_Y^2}{\theta_Y} + A_Y - 4\frac{\mu_{0300}}{\sigma_Y^2\mu_Y}\right]} \tag{2.2}$$

Motivated by the work of Misra et al (2017), we have proposed the following Modified Hansen Hurwitz estimator of finite population mean in the simultaneous presence of both non- response as well as measurement error.

$$\bar{y}_{HM} = \frac{1}{n}\left[n_1\bar{y}_{n_1} + n_2\left\{\bar{y}_{h_2} + k\left(\bar{y}_{h_2} - \frac{\hat{\sigma}_{y_{h_2}}}{C_{y_{n_2}}}\right)\right\}\right] \tag{2.3}$$

where $\bar{y}_{n_1} = \frac{1}{n_1}\sum_{i=1}^{n_1} y_i$ and $\bar{y}_{h_2} = \frac{1}{h_2}\sum_{i=1}^{h_2} y_i$

Here we have considered that sub sampled part h_2 of non-responding class n_2 contains measurement errors.

We have used the following approximations for calculation of Bias and Variance of proposed estimator.

$$\bar{y}_{h_2} = \bar{y}_{n_2}\left(1+e_o\right)$$

$$\hat{\sigma}^2_{y_{h_2}} = \sigma^2_{y_{n_2}}\left(1+e_1\right)$$

Where

$$E\left(e_o\right) = E\left(e_1\right) = 0,\ E\left(e_o^2\right) = \frac{C^2_{y_{n_2}}}{n_2\theta_{y_{n_2}}},\ \text{where } \theta_{y_{n_2}} = \frac{\sigma^2_{y_{n_2}}}{\sigma^2_{y_{n_2}} + \sigma^2_{u_{n_2}}},$$

$$E\left(e_1^2\right) = \frac{A_{y_{n_2}}}{n_2},$$

$$A_{Y_{N_2}} = \gamma_{Y_{N_2}} + \gamma_{u_{N_2}}\frac{\sigma^4_{u_{N_2}}}{\sigma^4_{Y_{N_2}}} + \left(1+\frac{\sigma^2_{u_{N_2}}}{\sigma^2_{Y_{N_2}}}\right)^2,\ E\left(e_oe_1\right) = \frac{\mu_3\left(y\right)}{n_2\sigma^2_{y_{n_2}}\bar{y}_{n_2}},$$

$$\gamma_{u_{n_2}} = \beta_2\left(u_{n_2}\right)-3,\quad \beta_2\left(u_{n_2}\right) = \frac{\mu_4\left(u_{n_2}\right)}{\mu_2^2\left(u_{n_2}\right)},\quad \mu_r\left(u\right) = E\left(u_{in_2} - \bar{u}_{n_2}\right)^r$$

The proposed estimator under the aforesaid approximation can be expressed as

$$\bar{y}_{HM} = \frac{1}{n}\left[n_1\bar{y}_{n_1} + n_2\left\{\bar{y}_{h_2} + k\left(\bar{y}_{h_2} - \frac{\hat{\sigma}_{y_{h_2}}}{C_{y_{n_2}}}\right)\right\}\right]$$

$$\bar{y}_{HM} = \frac{1}{n}\left[n_1\bar{y}_{n_1} + n_2\left\{\bar{y}_{n_2}\left(1+e_o\right) + k\left(\bar{y}_{n_2}\left(1+e_o\right) - \frac{\sigma_{yn_2}\left(1+e_1\right)^{\frac{1}{2}}}{C_{y_{n_2}}}\right)\right\}\right]$$

$$\bar{y}_{HM}=\frac{1}{n}\left[n_1\bar{y}_{n_1}+n_2\left\{\bar{y}_{n_2}+\bar{y}_{n_2}e_o+\frac{k}{C_{y_{n_2}}}\left(\bar{y}_{n_2}\left(1+e_o\right)C_{y_{n_2}}-\sigma_{y_{n_2}}\left(1+e_1\right)^{\frac{1}{2}}\right)\right\}\right]$$

$$\bar{y}_{HM}=\frac{1}{n}\left[n_1\bar{y}_{n_1}+n_2\left\{\bar{y}_{n_2}+\bar{y}_{n_2}e_o+\frac{k}{C_{y_{n_2}}}\sigma_{n_2}\left(\left(1+e_o\right)-\left(1+e_1\right)^{\frac{1}{2}}\right)\right\}\right]$$

$$\bar{y}_{HM}=\frac{1}{n}\left[n_1\bar{y}_{n_1}+n_2\left\{\bar{y}_{n_2}+\bar{y}_{n_2}e_o+k\bar{y}_{n_2}\left(\left(1+e_o\right)-\left(1+e_1\right)^{\frac{1}{2}}\right)\right\}\right]$$

$$\bar{y}_{HM}=\frac{1}{n}\left[n_1\bar{y}_{n_1}+n_2\bar{y}_{n_2}\left\{1+e_o+k\left(e_o-\tfrac{1}{2}e_1+\tfrac{1}{8}e_1^2\right)\right\}\right]$$

Taking expectations on both sides, where E_1, E_2 and E_3 represents the following expectations-

E_3 represents the expectation of total possible samples having their size h_2 taken from a samples of size m_2, E_2 represents expectation of all possible samples of size n_2 drawn from N_2,

E_1 stands for the expectation of total possible samples having their sizes n_1 taken from N_1.

$$E\left(\bar{y}_{HM}\right)=E_1E_2E_3\left[\frac{n_1}{n}\bar{y}_{n_1}+\frac{n_2}{n}\bar{y}_{n_2}\left\{1+e_o+k\left(e_o-\tfrac{1}{2}e_1+\tfrac{1}{8}e_1^2\right)\right\}\right]$$

$$E\left(\bar{y}_{HM}\right)=E_1E_2\left[\frac{n_1}{n}\bar{y}_{n_1}+\frac{n_2}{n}\bar{y}_{n_2}\left\{1+E_3\left(e_o\right)+k\left(E_3\left(e_o\right)-\tfrac{1}{2}E_3\left(e_1\right)+\tfrac{1}{8}E_3\left(e_1^2\right)\right)\right\}\right]$$

$$E\left(\bar{y}_{HM}\right)=E_1E_2\left[\frac{n_1}{n}\bar{y}_{n_1}+\frac{n_2}{n}\bar{y}_{n_2}\left\{1+k\tfrac{1}{8}E_3\left(e_1^2\right)\right\}\right]$$

$$E\left(\bar{y}_{HM}\right)=E_1E_2\left[\frac{n_1}{n}\bar{y}_{n_1}+\frac{n_2}{n}\bar{y}_{n_2}\left\{1+k\frac{1}{8n_2}\left(\gamma_{y_{n_2}}+\gamma_{u_{n_2}}\frac{\sigma_{u_{n_2}}^4}{\sigma_{y_{n_2}}^4}+\left(1+\frac{\sigma_{u_{n_2}}^2}{\sigma_{y_{n_2}}^2}\right)^2\right)\right\}\right]$$

$$E\left(\bar{y}_{HM}\right)=E_1\left(\frac{n_1}{n}\bar{y}_{n_1}\right)+E_1\left[E_2\left(\frac{n_2}{n}\bar{y}_{n_2}\left\{1+k\frac{1}{8n_2}\left(\gamma_{y_{n_2}}+\gamma_{u_{n_2}}\frac{\sigma^4_{u_{n_2}}}{\sigma^4_{y_{n_2}}}+\left(1+\frac{\sigma^2_{u_{n_2}}}{\sigma^2_{y_{n_2}}}\right)^2\right)\right\}\right)\right]$$

$$E\left(\bar{y}_{HM}\right)=E_1\left(\frac{n_1}{n}\bar{y}_{n_1}\right)+E_1\left[\frac{N_2}{N}\bar{Y}_{N_2}\left\{1+k\frac{1}{8N_2}\left(\gamma_{Y_{N_2}}+\gamma_{u_{N_2}}\frac{\sigma^4_{u_{N_2}}}{\sigma^4_{Y_{N_2}}}+\left(1+\frac{\sigma^2_{u_{N_2}}}{\sigma^2_{Y_{N_2}}}\right)^2\right)\right\}\right]$$

$$E\left(\bar{y}_{HM}\right)=\frac{N_1}{N}\bar{Y}_{N_1}+\frac{N_2}{N}\bar{Y}_{N_2}\left\{1+k\frac{1}{8N_2}\left(\gamma_{Y_{N_2}}+\gamma_{u_{N_2}}\frac{\sigma^4_{u_{N_2}}}{\sigma^4_{Y_{N_2}}}+\left(1+\frac{\sigma^2_{u_{N_2}}}{\sigma^2_{Y_{N_2}}}\right)^2\right)\right\}$$

$$Bias\left(\bar{y}_{HM}\right)=E\left(\bar{y}_{HM}\right)-\bar{Y}$$

$$Bias\left(\bar{y}_{HM}\right)=\frac{k}{8N}\bar{Y}_{N_2}\left(\gamma_{Y_{N_2}}+\gamma_{u_{N_2}}\frac{\sigma^4_{u_{N_2}}}{\sigma^4_{Y_{N_2}}}+\left(1+\frac{\sigma^2_{u_{N_2}}}{\sigma^2_{Y_{N_2}}}\right)^2\right)$$

$$k_{opt}=\frac{\dfrac{\mu_3\left(Y_{N_2}\right)}{\sigma^2_{Y_{N_2}}}-2\dfrac{C^2_{Y_{N_2}}}{\theta_{Y_{N_2}}}}{2\dfrac{C^2_{Y_{N_2}}}{\theta_{Y_{N_2}}}+\dfrac{A_{Y_{N_2}}}{2}-\dfrac{2\mu_3\left(Y_{N_2}\right)}{\sigma^2_{Y_{N_2}}\bar{Y}_{N_2}}},$$

After putting the value of K_{opt} theBias is

$$Bias\left(\bar{y}_{HM}\right)_{opt}=\frac{\dfrac{\mu_3\left(Y_{N_2}\right)}{\sigma^2_{Y_{N_2}}}-2\dfrac{C^2_{Y_{N_2}}}{\theta_{Y_{N_2}}}}{8N\left(2\dfrac{C^2_{Y_{N_2}}}{\theta_{Y_{N_2}}}+\dfrac{A_{Y_{N_2}}}{2}-\dfrac{2\mu_3\left(Y_{N_2}\right)}{\sigma^2_{Y_{N_2}}\bar{Y}_{N_2}}\right)}\bar{Y}_{N_2}\left(\gamma_{Y_{N_2}}+\gamma_{u_{N_2}}\frac{\sigma^4_{u_{N_2}}}{\sigma^4_{Y_{N_2}}}+\left(1+\frac{\sigma^2_{u_{N_2}}}{\sigma^2_{Y_{N_2}}}\right)^2\right)$$

$$Bias\left(\bar{y}_{HM}\right)_{opt}=\frac{\dfrac{\mu_3\left(Y_{N_2}\right)}{\sigma^2_{Y_{N_2}}}-2\dfrac{C^2_{Y_{N_2}}}{\theta_{Y_{N_2}}}}{8N\left(2\dfrac{C^2_{Y_{N_2}}}{\theta_{Y_{N_2}}}+\dfrac{A_{Y_{N_2}}}{2}-\dfrac{2\mu_3\left(Y_{N_2}\right)}{\sigma^2_{Y_{N_2}}\bar{Y}_{N_2}}\right)}\bar{Y}_{N_2}A_{Y_{N_2}}$$

$$Bias\left(\bar{y}_{HM}\right)_{opt}=\frac{\dfrac{\mu_3\left(Y_{N_2}\right)\theta_{Y_{N_2}}-2C^2_{Y_{N_2}}\sigma^2_{Y_{N_2}}}{\sigma^2_{Y_{N_2}}\theta_{Y_{N_2}}}}{8N\left(\dfrac{4C^2_{Y_{N_2}}\sigma^2_{Y_{N_2}}\bar{Y}_{N_2}+A_{Y_{N_2}}\sigma^2_{Y_{N_2}}\bar{Y}_{N_2}\theta_{Y_{N_2}}-4\mu_3\left(Y_{N_2}\right)\theta_{Y_{N_2}}}{2\sigma^2_{Y_{N_2}}\bar{Y}_{N_2}\theta_{Y_{N_2}}}\right)}\bar{Y}_{N_2}A_{Y_{N_2}}$$

$$Bias\left(\bar{y}_{HM}\right)_{opt}=\frac{\mu_3\left(Y_{N_2}\right)\theta_{Y_{N_2}}-2C^2_{Y_{N_2}}\sigma^2_{Y_{N_2}}}{4N\left(4C^2_{Y_{N_2}}\sigma^2_{Y_{N_2}}\bar{Y}_{N_2}+A_{Y_{N_2}}\sigma^2_{Y_{N_2}}\bar{Y}_{N_2}\theta_{Y_{N_2}}-4\mu_3\left(Y_{N_2}\right)\theta_{Y_{N_2}}\right)}\bar{Y}^2_{N_2}A_{Y_{N_2}}$$

3. Variance of Proposed Estimator

The variance of proposed estimator can be partitioned into two parts. We will find the variance for both parts separately and then combined both variance together.

$$V\left(\bar{y}_{HM}\right)=E\left(V\left(\bar{y}_{HM}/n_1,n_2\right)\right)+V\left(E\left(\bar{y}_{HM}/n_1,n_2\right)\right)$$

The variance of first part can be obtained as

$$\bar{y}_{HM}=\frac{1}{n}\left[n_1\bar{y}_{n_1}+n_2\left\{\bar{y}_{h_2}+k\left(\bar{y}_{h_2}-\frac{\hat{\sigma}_{y_{h_2}}}{C_{y_{n_2}}}\right)\right\}\right]$$

$$\bar{y}_{HM}=\frac{1}{n}\left[n_1\bar{y}_{n_1}+n_2\left\{\bar{y}_{n_2}\left(1+e_o\right)+k\left(\bar{y}_{n_2}\left(1+e_o\right)-\frac{\sigma_{y_{n_2}}\left(1+e_1\right)^{\frac{1}{2}}}{C_{y_{n_2}}}\right)\right\}\right]$$

$$\bar{y}_{HM}=\frac{1}{n}\left[n_1\bar{y}_{n_1}+n_2\left\{\bar{y}_{n_2}+\bar{y}_{n_2}e_o+\frac{k}{C_{y_{n_2}}}\left(\bar{y}_{n_2}\left(1+e_o\right)C_{y_{n_2}}-\sigma_{y_{n_2}}\left(1+e_1\right)^{\frac{1}{2}}\right)\right\}\right]$$

$$\bar{y}_{HM} = \frac{1}{n}\left[n_1 \bar{y}_{n_1} + n_2 \left\{ \bar{y}_{n_2} + \bar{y}_{n_2} e_o + \frac{k}{C_{y_{n_2}}} \sigma_{y_{n_2}} \left((1+e_o) - (1+e_1)^{\frac{1}{2}} \right) \right\} \right]$$

$$\bar{y}_{HM} = \frac{1}{n}\left[n_1 \bar{y}_{n_1} + n_2 \left\{ \bar{y}_{n_2} + \bar{y}_{n_2} e_o + k\bar{y}_{n_2} \left((1+e_o) - (1+e_1)^{\frac{1}{2}} \right) \right\} \right]$$

$$\bar{y}_{HM} = \frac{1}{n}\left[n_1 \bar{y}_{n_1} + n_2 \bar{y}_{n_2} \left\{ 1 + e_o + k\left(e_o - \tfrac{1}{2} e_1 + \tfrac{1}{8} e_1^2 \right) \right\} \right]$$

$$\bar{y}_{HM} = \bar{y} + \frac{n_2}{n} \bar{y}_{n_2} \left\{ e_o + k\left(e_o - \tfrac{1}{2} e_1 + \tfrac{1}{8} e_1^2 \right) \right\}$$

$$\bar{y}_{HM} - \bar{y} = \frac{n_2}{n} \bar{y}_{n_2} \left\{ e_o + k\left(e_o - \tfrac{1}{2} e_1 + \tfrac{1}{8} e_1^2 \right) \right\}$$

$$\left(\bar{y}_{HM} - \bar{y} \right)^2 = \left(\frac{n_2}{n} \bar{y}_{n_2} \right)^2 \left\{ e_o + k\left(e_o - \tfrac{1}{2} e_1 + \tfrac{1}{8} e_1^2 \right) \right\}^2$$

$$\left(\bar{y}_{HM} - \bar{y} \right)^2 = \left(\frac{n_2}{n} \bar{y}_{n_2} \right)^2 \left\{ e_o^2 + k^2\left(e_o^2 + \tfrac{1}{4} e_1^2 - e_o e_1 \right) + 2ke_o^2 - ke_o e_1 \right\}$$

$$E\left(\bar{y}_{HM} - \bar{y} \right)^2 = \left(\frac{n_2}{n} \bar{y}_{n_2} \right)^2 \left\{ (1+k)^2 E\left(e_o^2 \right) + \tfrac{1}{4} k^2 E\left(e_1^2 \right) - k(k+1) E\left(e_0 e_1 \right) \right\}$$

Putting the values of $E(e_o e_1)$, $E\left(e_o^2\right)$, $E\left(e_1^2\right)$

$$E\left(\bar{y}_{HM} - \bar{y} \right)^2 = V\left(\bar{y}_{HM} \right) = \left(\frac{n_2}{n} \bar{y}_{n_2} \right)^2 \left\{ (1+k)^2 \frac{C_{y_{n_2}}^2}{n_2 \theta_{y_{n_2}}} + \tfrac{1}{4} k^2 \frac{A_{y_{n_2}}}{n_2} - k(k+1) \frac{\mu_{3(y)}}{n_2 \sigma_{y_{n_2}}^2 \bar{y}_{n_2}} \right\}$$

$$E\left(\bar{y}_{HM} - \bar{y} \right)^2 = V\left(\bar{y}_{HM} \right) = \frac{n_2}{n^2} \bar{y}_{n_2}^2 \left\{ (1+k)^2 \frac{C_{y_{n_2}}^2}{\theta_{y_{n_2}}} + \tfrac{1}{4} k^2 A_{y_{n_2}} - k(k+1) \frac{\mu_{3(y)}}{\sigma_{y_{n_2}}^2 \bar{y}_{n_2}} \right\}$$

Now taking expectation

$$E\left(V\left(\bar{y}_{HM} \right) \right) = \frac{N_2}{nN} \bar{Y}_{N_2}^2 \left\{ (1+k)^2 \frac{C_{Y_{N_2}}^2}{\theta_{Y_{N_2}}} + \tfrac{1}{4} k^2 A_{Y_{N_2}} - k(k+1) \frac{\mu_{3(Y)}}{\sigma_{Y_{N_2}}^2 \bar{Y}_{N_2}} \right\}$$

Now second term of variance

$$\bar{y}_{HM} = \frac{1}{n}\left[n_1\bar{y}_{n_1} + n_2\left\{ \bar{y}_{h_2} + k\left(\bar{y}_{h_2} - \frac{\hat{\sigma}_{y_{h_2}}}{C_{y_{n_2}}} \right)\right\}\right]$$

$$\bar{y}_{HM} = \frac{1}{n}\left[n_1\bar{y}_{n_1} + n_2\left\{ \bar{y}_{n_2}\left(1+e_o\right) + k\left(\bar{y}_{n_2}\left(1+e_o\right) - \frac{\sigma_{y_{n_2}}\left(1+e_1\right)^{\frac{1}{2}}}{C_{y_{n_2}}} \right)\right\}\right]$$

$$\bar{y}_{HM} = \frac{1}{n}\left[n_1\bar{y}_{n_1} + n_2\left\{ \bar{y}_{n_2} + \bar{y}_{n_2}e_o + \frac{k}{C_{y_{n_2}}}\left(\bar{y}_{n_2}\left(1+e_o\right)C_{y_{n_2}} - \sigma_{y_{n_2}}\left(1+e_1\right)^{\frac{1}{2}} \right)\right\}\right]$$

$$\bar{y}_{HM} = \frac{1}{n}\left[n_1\bar{y}_{n_1} + n_2\left\{ \bar{y}_{n_2} + \bar{y}_{n_2}e_o + \frac{k}{C_{y_{n_2}}}\sigma_{y_{n_2}}\left(\left(1+e_o\right) - \left(1+e_1\right)^{\frac{1}{2}} \right)\right\}\right]$$

$$\bar{y}_{HM} = \frac{1}{n}\left[n_1\bar{y}_{n_1} + n_2\left\{ \bar{y}_{n_2} + \bar{y}_{n_2}e_o + k\bar{y}_{n_2}\left(\left(1+e_o\right) - \left(1+e_1\right)^{\frac{1}{2}} \right)\right\}\right]$$

$$\bar{y}_{HM} = \frac{1}{n}\left[n_1\bar{y}_{n_1} + n_2\bar{y}_{n_2}\left\{ 1 + e_o + k\left(e_o - \tfrac{1}{2}e_1 + \tfrac{1}{8}e_1^2 \right)\right\}\right]$$

$$\bar{y}_{HM} = \bar{y} + \frac{n_2}{n}\bar{y}_{n_2}\left\{ e_o + k\left(e_o - \tfrac{1}{2}e_1 + \tfrac{1}{8}e_1^2 \right)\right\}$$

E stands for the expectation of total possible samples having their sizes h_2 taken from the sample having size n_2.

$$E\left(\bar{y}_{HM}\right) = \bar{y} + \frac{n_2}{n}\bar{y}_{n_2}\left\{ E\left(e_o\right) + k\left(E\left(e_o\right) - \tfrac{1}{2}E\left(e_1\right) + \tfrac{1}{8}E\left(e_1^2\right) \right)\right\}$$

$$= \bar{y} + \frac{n_2}{n}\bar{y}_{n_2}\tfrac{1}{8}E\left(e_1^2\right)$$

$$E\left(\bar{y}_{HM}\right) = \bar{y} + \frac{n_2}{n}\bar{y}_{n_2}\frac{1}{8n_2}\left(\gamma_{y_{n_2}} + \gamma_{u_{n_2}}\frac{\sigma_{u_{n_2}}^4}{\sigma_{y_{n_2}}^4} + \left(1 + \frac{\sigma_{u_{n_2}}^2}{\sigma_{y_{n_2}}^2} \right)^2 \right)$$

Nowthe variance can be written as

$$V\left(E\left(\bar{y}_{HM}\right)\right)=V\left(\bar{y}\right)+V\left(\frac{n_2}{n}\bar{y}_{n_2}\frac{1}{8n_2}\left(\gamma_{y_{n_2}}+\gamma_{u_{n_2}}\frac{\sigma^4_{u_{n_2}}}{\sigma^4_{y_{n_2}}}+\left(1+\frac{\sigma^2_{u_{n_2}}}{\sigma^2_{y_{n_2}}}\right)^2\right)\right)$$

$$V\left(E\left(\bar{y}_{HM}\right)\right)=V\left(\bar{y}\right)+V\left(\frac{n_2}{n}\bar{y}_{n_2}\frac{1}{8n_2}A_{y_{n_2}}\right)$$

Ignoring $A_{y_{n_2}}$ due to higher order terms:

$$V\left(E\left(\bar{y}_{HM}\right)\right)=\left(\frac{1}{n}-\frac{1}{N}\right)S^2+\left(\frac{1}{8N}\right)^2\left(\frac{1}{n_2}-\frac{1}{N_2}\right)S^2_{N_2}$$

where $S^2=\frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i-\bar{Y}\right)^2$

Total Variance

$$MSE\left(\bar{y}_{HM}\right)=\left(\frac{1}{n}-\frac{1}{N}\right)S^2+\left(\frac{1}{8N}\right)^2\left(\frac{1}{n_2}-\frac{1}{N_2}\right)S^2_{N_2}$$
$$+\frac{N_2}{nN}\bar{Y}^2_{N_2}\left\{(1+k)^2\frac{C^2_{Y_{N_2}}}{\theta_{Y_{N_2}}}+\frac{1}{4}k^2A_{Y_{N_2}}-k\left(k+1\right)\frac{\mu_{3(Y)}}{\sigma^2_{Y_{N_2}}\bar{Y}_{N_2}}\right\}$$

The optimum value of k for which MSE of proposed estimator minimum is given as

$$k_{opt}=\frac{\frac{\mu_3\left(Y_{N_2}\right)}{\sigma^2_{Y_{N_2}}}-2\frac{C^2_{Y_{N_2}}}{\theta_{Y_{N_2}}}}{2\frac{C^2_{Y_{N_2}}}{\theta_{Y_{N_2}}}+\frac{A_{Y_{N_2}}}{2}-\frac{2\mu_3\left(Y_{N_2}\right)}{\sigma^2_{Y_{N_2}}\bar{Y}_{N_2}}},$$

After putting the optimum value of the characterizing scalar k, the minimum mean squared error of proposed estimator is calculated as,

$$MSE\left(\bar{y}_{HM}\right)_{opt}=\left(\frac{1}{n}-\frac{1}{N}\right)S^2+\left(\frac{1}{8N}\right)^2\left(\frac{1}{n_2}-\frac{1}{N_2}\right)S^2_{N_2}$$
$$+\frac{N_2}{nN}\bar{Y}^2_{N_2}\left\{\frac{2BD^2\bar{Y}^2_{N_2}+16AB\left(A-D\bar{Y}_{N_2}\right)+B^2D\bar{Y}_{N_2}-4A^3}{E^2}\right\} \quad (3.1)$$

where

$$A=\frac{\mu_3\left(Y_{N_2}\right)}{\sigma^2_{Y_{N_2}}},\quad B=\frac{C^2_{Y_{N_2}}}{\theta_{Y_{N_2}}},\quad D=\frac{A_{Y_{N_2}}}{2},\quad E=\left[\bar{Y}_{N_2}\left(2B+C\right)-2A\right]$$

4. Constraint Problem

We have assumed that population coefficient of variation is known to us from the pilot surveys and we using this information in our estimation procedure.

The MSE of proposed estimator is given as

$$MSE\left(\bar{y}_{HM}\right)=\left(\frac{1}{n}-\frac{1}{N}\right)S^2+\left(\frac{1}{8N}\right)^2\left(\frac{1}{n_2}-\frac{1}{N_2}\right)S^2_{N_2}$$
$$+\frac{N_2}{nN}\bar{Y}^2_{N_2}\left\{\left(1+k\right)^2\frac{C^2_{Y_{N_2}}}{\theta_{Y_{N_2}}}+\frac{1}{4}k^2A_{Y_{N_2}}-k\left(k+1\right)\frac{\mu_{3(Y)}}{\sigma^2_{Y_{N_2}}\bar{Y}_{N_2}}\right\}$$

The population coefficient of variation $\frac{s_{y_{h_2}}}{\bar{y}_{h_2}}k=C_{y_{n_2}}$ is known.

So our problem is to minimize V($\bar{y}_{HM}$), provided that population coefficient of variation C_Y is known.

Now the optimization problem is

$$\phi=V\left(\bar{y}_{HM}\right)-\lambda\left(\frac{s_{y_{h_2}}}{\bar{y}_{h_2}}k-C_{y_{n_2}}\right)$$

where λ is a Lagrange multiplier.

$$\frac{d\phi}{dk}=\frac{N_2}{nN}\bar{Y}^2_{N_2}\left\{2\left(1+k\right)\frac{C^2_{Y_{N_2}}}{\theta_{Y_{N_2}}}+\frac{1}{2}kA_{Y_{N_2}}-2k\frac{\mu_{3(Y)}}{\sigma^2_{Y_{N_2}}\bar{Y}_{N_2}}-\frac{\mu_{3(Y)}}{\sigma^2_{Y_{N_2}}\bar{Y}_{N_2}}\right\}-\lambda\frac{s_{y_{h_2}}}{\bar{y}_{h_2}}$$

Equating to zero

$$k\frac{N_2}{nN}\overline{Y}_{N_2}^2\left\{2\frac{C_{Y_{N_2}}^2}{\theta_{Y_{N_2}}}+\frac{1}{2}kA_{Y_{N_2}}-2\frac{\mu_3(Y)}{\sigma_{Y_{N_2}}^2\overline{Y}_{N_2}}\right\}-\frac{N_2}{nN}\overline{Y}_{N_2}^2\left\{2\frac{C_{Y_{N_2}}^2}{\theta_{Y_{N_2}}}-\frac{\mu_3(Y)}{\sigma_{Y_{N_2}}^2\overline{Y}_{N_2}}\right\}=\lambda\frac{s_{y_{h_2}}}{\overline{y}_{h_2}}$$

$$k=\frac{\lambda\frac{s_{y_{h_2}}}{\overline{y}_{h_2}}-\frac{N_2}{nN}\overline{Y}_{N_2}^2\left\{2\frac{C_{Y_{N_2}}^2}{\theta_{Y_{N_2}}}-\frac{\mu_3(Y)}{\sigma_{Y_{N_2}}^2\overline{Y}_{N_2}}\right\}}{\frac{N_2}{nN}\overline{Y}_{N_2}^2\left\{2\frac{C_{Y_{N_2}}^2}{\theta_{Y_{N_2}}}+\frac{1}{2}kA_{Y_{N_2}}-2\frac{\mu_3(Y)}{\sigma_{Y_{N_2}}^2\overline{Y}_{N_2}}\right\}}$$

Using the value of k in constraint

$$s_{y_{h_2}}\left[\lambda\frac{s_{y_{h_2}}}{\overline{y}_{h_2}}-\frac{N_2}{nN}\overline{Y}_{N_2}^2\left\{2\frac{C_{Y_{N_2}}^2}{\theta_{Y_{N_2}}}-\frac{\mu_3(Y)}{\sigma_{Y_{N_2}}^2\overline{Y}_{N_2}}\right\}\right]$$

$$=C_{y_{n_2}}\overline{y}_{h_2}\frac{N_2}{nN}\overline{Y}_{N_2}^2\left\{2\frac{C_{Y_{N_2}}^2}{\theta_{Y_{N_2}}}+\frac{1}{2}kA_{Y_{N_2}}-2\frac{\mu_3(Y)}{\sigma_{Y_{N_2}}^2\overline{Y}_{N_2}}\right\}$$

$$\lambda s_{y_{h_2}}^2=s_{y_{h_2}}\frac{N_2}{nN}\overline{Y}_{N_2}^2\left\{2\frac{C_{Y_{N_2}}^2}{\theta_{Y_{N_2}}}-\frac{\mu_3(Y)}{\sigma_{Y_{N_2}}^2\overline{Y}_{N_2}}\right\}+C_{y_{n_2}}\overline{y}_{h_2}\frac{N_2}{nN}\overline{Y}_{N_2}^2\left\{2\frac{C_{Y_{N_2}}^2}{\theta_{Y_{N_2}}}+\frac{1}{2}kA_{Y_{N_2}}-2\frac{\mu_3(Y)}{\sigma_{Y_{N_2}}^2\overline{Y}_{N_2}}\right\}$$

Putting the value of $\lambda\frac{s_{y_{h_2}}}{\overline{y}_{h_2}}$ in k

$$k=\frac{\frac{N_2}{nN}\overline{Y}_{N_2}^2\left\{2\frac{C_{Y_{N_2}}^2}{\theta_{Y_{N_2}}}-\frac{\mu_3(Y)}{\sigma_{Y_{N_2}}^2\overline{Y}_{N_2}}\right\}\frac{C_{y_{n_2}}}{s_{y_{h_2}}}\overline{y}_{h_2}\frac{N_2}{nN}\overline{Y}_{N_2}^2\left\{2\frac{C_{Y_{N_2}}^2}{\theta_{Y_{N_2}}}+\frac{1}{2}kA_{Y_{N_2}}-2\frac{\mu_3(Y)}{\sigma_{Y_{N_2}}^2\overline{Y}_{N_2}}\right\}-\frac{N_2}{nN}\overline{Y}_{N_2}^2\left\{2\frac{C_{Y_{N_2}}^2}{\theta_{Y_{N_2}}}-\frac{\mu_3(Y)}{\sigma_{Y_{N_2}}^2\overline{Y}_{N_2}}\right\}}{\frac{N_2}{nN}\overline{Y}_{N_2}^2\left\{2\frac{C_{Y_{N_2}}^2}{\theta_{Y_{N_2}}}+\frac{1}{2}kA_{Y_{N_2}}-2\frac{\mu_3(Y)}{\sigma_{Y_{N_2}}^2\overline{Y}_{N_2}}\right\}}$$

$$k = C_{y_{n2}} \frac{\bar{y}_{h_2}}{s_{y_{h_2}}}$$

5. Theoretical Efficiency Comparison

We have made a theoretical efficiency comparison between the proposed estimator and usual Hansen Hurvitz estimator.

Variance of usual Hansen Hurvitz estimator is given as `

$$V(\bar{y}_w) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^{\,2} + \frac{f-1}{n}\frac{N_2}{N} S_{y_2}^2$$

MSE of proposed estimator is given by

$$MSE(\bar{y}_{HM})_{opt} = \left(\frac{1}{n} - \frac{1}{N}\right) S^2 + \left(\frac{1}{8N}\right)^2 \left(\frac{1}{n_2} - \frac{1}{N_2}\right) S_{N_2}^2 + \frac{N_2}{nN} \bar{Y}_{N_2}^2 \left\{ \frac{2BD^2\bar{Y}_{N_2}^2 + 16AB\left(A - D\bar{Y}_{N_2}\right) + B^2 D\bar{Y}_{N_2} - 4A^3}{E^2} \right\}$$

The suggested estimator will perform better than the usual Hansen- Hurvitz estimator if,

$MSE(\bar{y}_{HM})_{opt} < MSE(\bar{y}_w)$

which gives optimality condition

$$\left(\frac{1}{8N}\right)^2 \left(\frac{1}{n_2} - \frac{1}{N_2}\right) S_{N_2}^2 + \frac{N_2}{nN} \bar{Y}_{N_2}^2 \left\{ \frac{2BD^2\bar{Y}_{N_2}^2 + 16AB\left(A - D\bar{Y}_{N_2}\right) + B^2 D\bar{Y}_{N_2} - 4A^3}{E^2} \right\} < \frac{f-1}{n}\frac{N_2}{N} S_{y_2}^2 \qquad (5.1)$$

6. A simulation Study

We demonstrate the performances of both estimators through simulation study by generating a sample from Normal distribution using R software. The considered population is very much relevant in various socio-economic situations.

Description of data is given as follows

$N = 5000$, $N_1 = 4000$, $N_2 = 1000$, $n = 500$, $n_1 = 300$, $n_2 = 200$, $h_2 = 50$.

Y=N (5000, 2, 4)

Variance (MSE) of Hansen Hurwitz estimator and proposed estimator for the situation, when the non-response as well as the measurement errors both are present simultaneously.

Hansen Hurwitz estimator $V(\bar{y}_{HH}) = 0.05880166$

Proposed estimator $MSE(\bar{y}_{HM})_{opt} = 0.03270853$

7. Conclusion

From the simulation study, it is evident that the proposed estimator performs better than the usual Hansen Hurvitz estimator in the simultaneous presence of non-response and measurement errors as it is more efficient than it. Therefore proposed estimator is recommended to survey practitioners to estimate population mean in the case when survey data is contaminated with non-response as well as measurement errors.

Acknowledgment

The authors are very much thankful to the anonymous referees and editor of JRSS for their useful suggestions which improved the earlier draft.

References

1. Cochran, W.G. (1963). Sampling Techniques, Second Edition, Wiley Eastern Private Limited, New Delhi.
2. Cochran, W.G.(1968). Errors of measurement in statistics, Technometrics, 10, p. 637-666.
3. Devi, M., Sisodia,B.V.S. and Azfer, M.(2016). Estimation of population mean in two-stage sampling under a random response mechanism in the presence of non-response, Inter. Jour. Agri. Statistics, 12(1), p. 27-34.
4. Foradori, G. T. (1961). Some non-response sampling theory for two stage designs, Institute of Statistics, North Carolina State College.
5. Hansen, M.H. and Hurwitz, W.N.(1946). The problem of non-response in sample surveys, Journal of American Statistical Association, 41, p. 517-529.
6. Khare, B. B., Jha, P. S. and Kumar, K. (2015). Estimation of population proportion using auxiliary character in the presence of nonresponse, Journal of Reliability and Statistical Studies, 8(2), p. 107-117.
7. Kumar,S. (2014). Variance estimation in presence of random non-response, Journal of Reliability and Statistical Studies, 7(1), p. 65-70.
8. Misra, S. and Yadav, D. K. (2015). Estimating population mean using known coefficient of variation under measurement errors, Statistics and Informatics in Agricultural Research, Excel India Publisher, New Delhi, p. 175-182.
9. Misra, S., Dipika and Yadav, D. K.(2016). Some improved estimators for estimating population variance in the presence of measurement errors, Journal of Statistics Applications & Probability, Natural Science Publishing, USA, 2 (5), p. 311-320.
10. Misra, S., Yadav, D.K. and Dipik. (2016). An efficient estimator for estimating population variance in presence of measurement errors, International Journal of Mathematics and its Applications, 4(2), p. 23-28.
11. Misra, S., Yadav D.K. and Dipika (2017). Estimation of population mean using auxiliary information in presence of measurement errors, International Journal of Engineering Sciences & Research Technology, 6(6), p. 499-502.
12. Okafor, F. C. (2001). Treatment of non-response in successive sampling, Statistica, 61(2), p.195-204.

13. Okafor, F. C. (2005). Sub-sampling the non-respondents in two-stage sampling over two successive occasions, Journal of the Indian Statistical Association, 43(1), p.33-49.
14. Shalabh (1997). Ratio method of estimation in the presence of measurement errors, Jour. Ind. Soc. Ag. Statistics, 1(2), p.150-155.
15. Singh, H. P. and Karpe, N. (2008). Ratio-product estimator for population mean in presence of measurement errors, Journal of Applied Statistical Sciences, 16, p.49–64.
16. Singh, H.P. and Karpe, N. (2009). A general procedure for estimating the general parameter using auxiliary information in presence of measurement errors, Communication of the Korean Statistical Society, 16(5), p.821-840.
17. Singh, H. P. and Karpe, N. (2010). Estimation of mean, ratio and product using auxiliary information in the presence of measurement errors in sample surveys, Journal of Statistical Theory and Practice, 4(1), p.111–13
18. Singh, H.P. and Kumar,S. (2011). Subsampling the non-respondents in cluster sampling on sampling on two successive occasions, Statistics in Transition-New Series, 12(1), p. 9-24.
19. Singh, R.K. and Naqvi, N. (2015). Estimation of population mean using auxiliary information in non-response, Journal of Reliability and Statistical Studies, 8(1), p.135-14.
20. Srivastava, A.K. and Shalabh (2001). Effect of measurement errors on the regression method of estimation in survey sampling, Journal of Statistical Research, 35, p.35-44.
21. Sud, C. and Srivastava, S. K. (2000). Estimation of population mean in repeat surveys in the presence of measurement errors, Journal of Indian Society of Agricultural Statistics, 53, p.125-133.
22. Sukhatme, P.V, Sukhatme, B.V. ,Sukhatme, S. and Ashok, C. (1984). Sampling Theory of Surveys with Applications, 3rd Ed., Iowa State University Press, Ams, Iowa(USA) and Indian Society of Agricultural Statistics, New Delhi.
23. Tripathi, T. P. and Khare, B.B. (1997). Estimation of mean vector in presence of nonresponse, Communications in Statistics-Theory and Methods, 26(9), p. 2255-2269.