

ISSN Online: 2456-1452
NAAS Rating (2026): 4.49
IJSAM 2026; 11(10): 30-32
www.mathsjournal.com
Received: 03-08-2026
Accepted: 25-09-2026
Published: 07-10-2026

Barasa Laban Sakwa
Department of Mathematics,
Actuarial Science Kenyatta
University, Nairobi, Kenya

Christopher Ouma Onyango
Department of Mathematics,
Actuarial Science Kenyatta
University, Nairobi, Kenya

An improved ratio estimator for population mean in the presence of non-response in two stage sampling using two auxiliary variables

Barasa Laban Sakwa¹, Christopher Ouma Onyango²

DOI: <https://www.doi.org/10.22271/math.2026.v11.i10a.2714>

ABSTRACT

Nonresponse is a common challenge during sample surveys that tend to interfere with efficiency of population mean estimators. Nonresponse occurs when selected units fail to provide the required information. If ignored the resulting estimator may be unbiased. The study proposes an improved class of ratio estimators for estimating the finite population mean using information from two auxiliary variables and two stage sampling where nonresponse occurs at the second stage of sampling. The estimator allows the two auxiliary variables to contribute simultaneously to the estimation of the population mean. Using first order Taylor approximation, expressions for bias and mean squared error (MSE) of the proposed estimators are derived. Theoretical efficiency conditions are established. Empirical comparison shows the proposed estimators outperform existing by Singh *et al.* (2018) with a percentage relative efficiency (PRE) of over 120%

Keywords: Ratio estimator, Double sampling, Nonresponse, Auxiliary variables, Mean squared error, Bias

1. Introduction

In survey sampling, auxiliary information is essential for improving precision and accuracy of estimators for population parameter of interest. Such information can be utilized both during sample selection and the estimation stage to enhance the design and yield more consistent estimates. For example, estimation becomes more precise when the auxiliary variable X is highly correlated with the study variable Y . In many practical survey, information on auxiliary variables may not be available for the entire population, necessitating two stage sampling.

Furthermore, nonresponse is unavoidable in survey relating to health, education and industry. When nonresponse occurs, the Hansen and Hurwitz (1946) subsampling technique is widely adopted. Singh *et al.* (2012) proposed estimators under two-stage sampling with nonresponse using one auxiliary variable, Ongoma *et al.* (2021) [1] considered nonresponse in both study and auxiliary, limited work has addressed the use of two auxiliary variables that are also subject to nonresponse at the second phase. This gap motivates the present study.

The objective of this paper is to propose a ratio estimator using two auxiliary variables under two stage sampling with nonresponse and derive its properties.

2. Methodology

2.1 Notation and sampling design

A finite population consisting N primary samples from $\beta_1, \beta_2, \dots, \beta_N$ clusters of sizes M_i for $i = 1, 2, \dots, N$. A sample of n clusters were considered from N clusters using SRSWOR as the ssu 's, assuming that m_i units provide the required information on the two auxiliary variables; Due to nonresponse, only m_{i1} out of the m_i units selected provide complete data of study variable and two auxiliary variables. Remaining m_{i2} units do not respond. Non response occurs only at the second stage and is observed both on study and auxiliary variables. By the use of Hansen and Hurwitz methods of taking a subsample of the units that fail to respond where another sample of h_{i2} units are taken out of the total m_{i2} that fail to respond in such a manner that $m_{i2} = k_i h_{i2}$ where $k_i > 1$. Let Y be study variables and X_1, X_2 be the two auxiliary variables. Therefore, the samples picked on the second phase will have a total of the responding (m_{i1}) and nonresponding (m_{i2}) that is, $m_i = m_{i1} + m_{i2}$. Let h_{i2} units be a fraction of non-responding group that respond after callback, then the total responding units are; $m_{i1} + h_{i2}$. with k_i as inverse sampling rate. For the responding units $W_i = \frac{m_{i1}}{m_i}$ as their cluster weight and $W_i = \frac{m_{i2}}{m_i}$ cluster weight for the non-responding units.

Corresponding Author:
Barasa Laban Sakwa
Department of
Mathematics, Actuarial
Science Kenyatta
University, Nairobi, Kenya

2.2 Proposed estimator

The proposed ratio-type estimator using two auxiliary variables is defined as:

$$\hat{Y}_i = \frac{\bar{y}_i}{\bar{x}_i \bar{z}_i} \bar{x}_i \bar{z}_i$$

with an assumption that $E(\bar{z}_{ij}) = E(\bar{z}_{ij}) = \bar{Z}_{ij}$ and $E(\bar{x}_i) = E(\bar{x}_i) = \bar{X}_{ij}$ and $E(\bar{y}_i) = \bar{Y}_{ij}$

To obtain MSE and Bias we will use the large sample approximation and Taylor series expansion.

2.3 Bias and mean squared error

define

$$\varepsilon_0 = \frac{\bar{y}_i - \bar{Y}_{ij}}{\bar{Y}_{ij}} \quad \text{therefore} \quad \bar{y}_i = \bar{Y}_{ij}(\varepsilon_0 + 1) \quad , \quad \varepsilon_1 = \frac{\bar{x}_i - \bar{X}_{ij}}{\bar{X}_{ij}} \quad \text{therefore}$$

$$\bar{x}_i = \bar{X}_{ij}(\varepsilon_1 + 1)$$

$$\varepsilon_2 = \frac{\bar{x}_i - \bar{X}_{ij}}{\bar{X}_{ij}} \quad \text{therefore} \quad \bar{x}_i = \bar{X}_{ij}(\varepsilon_2 + 1), \quad \varepsilon_3 = \frac{\bar{z}_i - \bar{Z}_{ij}}{\bar{Z}_{ij}} \quad \text{therefore}$$

$$\bar{z}_i = \bar{Z}_{ij}(\varepsilon_3 + 1)$$

$$\varepsilon_4 = \frac{\bar{z}_i - \bar{Z}_{ij}}{\bar{Z}_{ij}} \quad \text{therefore} \quad \bar{z}_i = \bar{Z}_{ij}(\varepsilon_4 + 1), \quad \text{Where}$$

$$E(\varepsilon_i) = 0 \text{ for } i = 1, 2, 3, 4$$

Expanding to the second order of approximation, the bias is

Bia

$$(\bar{Y}_i) = \left(\frac{1}{m_i} - \frac{1}{N} \right) (C_x^2 + C_z^2 + \rho C_x C_z - \rho C_x C_y - C_y C_z) + \left(\frac{1}{m_i} - \frac{1}{N} \right) (\rho C_x^2 + \rho C_z^2 + \rho C_x C_z - \rho C_y C_z - \rho C_x C_y) + \left(\frac{k-1}{m_i} \right) C_x^2 +$$

We then find MSE by squaring both sides of the estimator and taking expectation neglecting powers that are >2 we have:

$$MSE(\hat{Y}_i) = \left(\frac{1}{m_i} - \frac{1}{N} \right) S_{R1} + \left(\frac{1}{m_i} - \frac{1}{m_i} \right) S_{R2} + \left(\frac{k-1}{m_i} \right) w_{i2} S_{R3}$$

Were

$$S_{R1} = S_y^2 + S_x^2 R_x^2 + S_z^2 R_z^2 + S_z^2 R_z^2 + 2\rho_{yz} S_y S_z R_z + 2\rho_{xz} S_x S_z R_x R_z - 2S_x^2 R_x^2 - \rho_{xz} S_x S_z R_x R_z - 2S_z^2 R_z^2$$

$$S_{R2} = \left(S_y^2 + S_x^2 R_x^2 + 2\rho_{xz} S_x S_z R_x R_z - 2\rho_{xy} S_x S_y R_x - 2\rho_{yz} S_y S_z R_z \right)$$

$$S_{R3} = \left(S_y^2 + S_x^2 R_x^2 + S_z^2 R_z^2 - 2\rho_{xy} S_x S_y R_x \right)$$

Table 1. The simulation results summary for Bias and MSE for one and two auxiliary variables

Estimator	Bias	Variance	Mse	%bias reduction vs naïve	% MSE reduction vs naïve
Naïve	0.0123	1.7591	1.7593	-	-
1-auxiliary ratio	-0.0011	0.9673	0.9673	75.9%	77.1%
2-auxiliary regression	0.0004	0.9599	0.9599	91.3%	84.1%

The proposed estimator has the lowest MSE

3.Results: Efficiency comparison

The following estimators were computed in each application:

Naïve estimator that uses only respondent data while ignoring auxiliary variables and nonresponse mechanism. One auxiliary ratio estimator that adjusts for the nonresponse while only making use of one auxiliary variable. Two auxiliary ratio estimators. The proposed estimator that listens to the nonresponse while in cooperating two auxiliary variables

The typical results obtained from 1000 replications, with the population mean 89.4 and a 30% MAR nonresponse were as per the table below.

4. Conclusion

The aim of this study was to come up with an estimator that incorporates two auxiliary variables considering two stage sampling with an element of nonresponse.

Non response led the Naïve estimator to decrease its bias but when we incorporated the application of two auxiliary variables the bias was reduced to as much as 91%. The use of two auxiliary variable as well lowered the MSE by 84% as compared to the Naive estimator.

The use of two auxiliary variables also lowered the MSE by 31% as compared to when using one auxiliary variable.

References

1. Ongoma Jackson, Alila David Anekeya, Okuto Erick. Optimal Allocation in Small Area Mean Estimation Using Stratified Random Sampling in the presence of non-response. *International Journal of Statistical Distributions and Application*. 2021;7(1):13-24.
2. Rajesh Singh, Manoj kr Chaudhary, Anil Prajapati. two-phase sampling in estimation of population mean in the presence of non-response. Columbus, USA: Educational publisher; 2014:28-41.
3. Alila David Anekeya, Ouma Christopher Onyango, Nyongesa Kennedy. Domain Mean Estimation Using Double Sampling with Non-Linear Cost Function in the Presence of Non-Response. *Science Journal of Applied Mathematics and Statistics*. 2018;6(1):28-42.
4. Nelson Kiprono Bii, Christopher Ouma Onyango, John Odhiambo. Estimating a finite population mean and random non-response in two stage cluster sampling with replacement. *Journal of physics. conf series*. 2018;1132:012071.

5. Monoj K Chaudhary, VK Singh. Setting the population mean in two-stage sampling with equal size clusters under non-response using auxiliary characteristic. Mathematical journal of interdisciplinary sciences. 2013;2(1):43-55.

How to cite this article: Sakwa BL, Onyango CO. An improved ratio estimator for population mean in the presence of non-response in two stage sampling using two auxiliary variables. *International Journal of Statistics and Applied Mathematics*. 2026;11(10):30-32. doi: <https://www.doi.org/10.22271/math.2026.v11.i10a.2714>