International Journal of Statistics and Applied Mathematics 2019; 4(6): 37-44

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2019; 4(6): 37-44

www.mathsjournal.com
Received: 16-09-2019
Accepted: 18-10-2019

Onyango O Ronald
School of Mathematics and Actuarial Science, Jaramogi Oginga Odinga University of Science and Technology, Bondo, Kenya

Oduor Brian
School of Mathematics and Actuarial Science, Jaramogi Oginga Odinga University of Science and Technology, Bondo, Kenya

Odundo Francis
School of Mathematics and Actuarial Science, Jaramogi Oginga Odinga University of Science and Technology, Bondo, Kenya

Corresponding Author:
Onyango O Ronald
School of Mathematics and Actuarial Science, Jaramogi Oginga Odinga University of Science and Technology, Bondo, Kenya

Optimal allocation in double sampling for stratification in the presence of nonresponse and measurement errors

Onyango O Ronald, Oduor Brian and Odundo Francis

Abstract
The present study addresses the problem of minimum cost and precision in the estimation of the population mean in the presence of nonresponse and measurement errors. It is assumed that both the survey variable and the auxiliary variable suffer from nonresponse and measurement errors in the second phase sample. A ratio, exponential ratio-ratio type, and exponential product-ratio type estimators of the population mean are proposed using the information on a single auxiliary variable. The expression of biases and mean square errors of the proposed estimators are obtained up to the first order of approximation. The cost of the survey is studied theoretically. The optimum stratum sample size and the inverse sampling rate are derived. It is noted that the size of the sample to be selected increases with the increase in the cost of the survey. A minimum mean square error is attained when the cost of the survey is high.

Keywords: Double sampling for stratification, population mean, nonresponse, and measurement error.

1. Introduction
1.1 Nonresponse and Measurement Errors
The presence of nonresponse and measurement errors in a survey contaminates data making analysis to result in under-estimated or over-estimated parameters. This leads to invalid statistical inferences that may have undesirable consequences in policymaking. In a survey, a researcher faces the problem of estimating the population mean and choosing the optimal stratum sample size that minimizes the variance for a specified cost in the presence of errors. In the literature authors who have studied nonresponse under the different sampling schemes include [1-5] and measurement errors include [6-8].

The precision of the statistic being estimated in a survey increases with the decrease in the variance. The variance of the statistic depends on the size of the strata which can be determined prior using the allocation method. The optimization technique is used in a survey to obtain a robust estimator of the population parameter for a fixed cost. Double sampling for stratification encounters a major drawback of determining the first phase and the second phase stratum sample sizes that give the desired precision for a specified cost.

The problem of optimal allocation in the estimation of the population mean using the auxiliary variable in the presence of errors is not addressed in the literature. The aim of the present study is to use the information on a single auxiliary variable to propose estimators of the population mean in the presence of nonresponse and measurement errors. The expression of biases and mean square errors of the proposed estimators are obtained. The cost of the survey is studied theoretically. The optimum sample sizes and the value of the inverse sampling rate are derived.

1.2 Sampling Procedure
In double sampling for stratification, a heterogeneous population of size N is considered. A first phase sample of size n' is drawn from the population using a simple random sampling without replacement design and the units classified into L homogeneous strata of size n'_h each. The auxiliary variable is studied in the first phase sample. A second phase random sample of

size n_h is drawn from the first phase sample and both the survey variable and the auxiliary variable are studied.

The h^{th} stratum first phase sample mean of the auxiliary variable is given as $\bar{x}'_h = \frac{1}{n'_h}\sum_{i=1}^{n'_h} x'_{hi}$. The second phase sample is divided into responding and nonresponding groups of sizes n_{1h} and n_{2h} respectively. A random sample of size $r_{2h} = \frac{n_{2h}}{k_{2h}}$, where $k_{2h} > 1$ is the inverse sampling rate, is drawn from the nonresponding group and used in the estimation of the population mean. Let (y_{hi}, x_{hi}) denote the survey variable and the auxiliary variable respectively. The population mean of the survey variable and the auxiliary variable in the h^{th} stratum is given as $\bar{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} Y_{hi}$ and $\bar{X}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} X_{hi}$ respectively. The population mean of the survey variable and the auxiliary variable for the nonresponding units in the h^{th} stratum is given as $\bar{Y}_{2h} = \frac{1}{N_{2h}}\sum_{i=1}^{N_{2h}} Y_{hi}$ and $\bar{X}_{2h} = \frac{1}{N_{2h}}\sum_{i=1}^{N_{2h}} X_{hi}$ respectively. The h^{th} stratum population variance of the survey variable is given as $S_{Yh}^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(y_{hi} - \bar{Y}_h)^2$ while that of the nonresponding units is given as $S_{Yh(2)}^2 = \frac{1}{N_{2h}-1}\sum_{i=1}^{N_{2h}}(y_{hi} - \bar{Y}_{2h})^2$. The h^{th} stratum population variance of the auxiliary variable is given as $S_{Xh}^2 = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(x_{hi} - \bar{X}_h)^2$ while that of the nonresponding units is given as $S_{Xh(2)}^2 = \frac{1}{N_{2h}-1}\sum_{i=1}^{N_{2h}}(x_{hi} - \bar{X}_{2h})^2$. The estimator of the population mean proposed by [9] in the presence of nonresponse is extended to double sampling for stratification and is defined as $\bar{y}_h^* = \frac{n_{1h}}{n_h}\bar{y}_{n1h} + \frac{n_{2h}}{n_h}\bar{y}_{r2h}$, where $\bar{y}_{n1h}$ and $\bar{y}_{r2h}$ are the sample mean of the responding units and the subsample mean of the nonresponding units respectively. The population mean of the auxiliary variable is given as $\bar{x}_h^* = \frac{n_{1h}}{n_h}\bar{x}_{n1h} + \frac{n_{2h}}{n_h}\bar{x}_{r2h}$, where $\bar{x}_{n1h}$ and $\bar{x}_{r2h}$ are the sample mean of the responding units and the subsample mean of the nonresponding units respectively.

Let the observed values of the auxiliary variable and the survey variable be (x_{hi}^*, y_{hi}^*) and their corresponding true values be (X_{hi}^*, Y_{hi}^*) respectively in the presence of measurement errors. Define the measurement errors associated with the survey variable as $U_{hi}^* = y_{hi}^* - Y_{hi}^*$ and those associated with the auxiliary variable as $V_{hi}^* = x_{hi}^* - X_{hi}^*$. The measurement errors are independent and uncorrelated. They are assumed to occur randomly in nature with mean zero. The variance of the measurement errors of the survey variable and the auxiliary variable for the responding units are S_{Uh}^2 and S_{Vh}^2 while for the nonresponding units are $S_{Uh(2)}^2$ and $S_{Vh(2)}^2$ respectively.

2. The Proposed Estimators of Population Mean

The study proposes the following estimators of the population mean in the presence of nonresponse and measurement errors on both the auxiliary variable and the survey variable in the second phase sample. The ratio estimator is defined as

$$T_R = \sum_{h=1}^{L} w_h \bar{y}_h^* \left(\frac{\bar{x}'_h}{\bar{x}_h^*}\right)$$

The exponential ratio-ratio type estimator is defined as

$$T_{ERR} = \sum_{h=1}^{L} w_h \bar{y}_h^* \left(\frac{\bar{x}'_h}{\bar{x}_h^*}\right) exp\left(\frac{\bar{x}'_h - \bar{x}_h^*}{\bar{x}'_h + \bar{x}_h^*}\right)$$

The exponential product-ratio type estimator is defined as

$$T_{EPR} = \sum_{h=1}^{L} w_h \bar{y}_h^* \left(\frac{\bar{x}'_h}{\bar{x}_h^*}\right) exp\left(\frac{\bar{x}_h^* - \bar{x}'_h}{\bar{x}_h^* + \bar{x}'_h}\right)$$

To obtain the expression of biases and mean square errors of the proposed estimators, let

$$\sigma_{Yh}^{*2} = \frac{1}{n_h}\sum_{i=1}^{n_h}(y_{hi}^* - \bar{Y}_h)^2$$

Multiply both sides by $\frac{1}{n_h}$ and introduce the square root to obtain

$$\frac{1}{\sqrt{n_h}}\sigma_{Yh}^* = \frac{1}{n_h}(y_{hi}^* - \bar{Y}_h) \quad (1)$$

Also, define

$$\sigma_{Uh}^{*2} = \frac{1}{n_h}\sum_{i=1}^{n_h}(y_{hi}^* - Y_{hi}^*)^2$$

Multiply both sides by $\frac{1}{n_h}$ and introduce the square root to obtain

$$\frac{1}{\sqrt{n_h}}\sigma_{Uh}^* = \frac{1}{n_h}\sum_{i=1}^{n_h}(y_{hi}^* - Y_{hi}^*) \quad (2)$$

Combine equation (1) and (2) to obtain

$$\frac{1}{\sqrt{n_h}}(\sigma_{Yh}^* + \sigma_{Uh}^*) = \frac{1}{n_h}\sum_{i=1}^{n_h}\{(y_{hi}^* - \bar{Y}_h) + (y_{hi}^* - Y_{hi}^*)\} \tag{3}$$

Let $\frac{1}{\sqrt{n_h}}(\sigma_{Yh}^* + \sigma_{Uh}^*) = \sigma_{Yh}$ and simplify equation (3) to obtain

$$\sigma_{Yh} = y_h^* - \bar{Y}_h \tag{4}$$

Similarly, the following can be obtained for the auxiliary variable

$$\sigma_{Xh} = x_h^* - \bar{X}_h \tag{5}$$

Since the first phase sample does not suffer from nonresponse and measurement errors,

$$\sigma_{X1h} = x_h' - \bar{X}_h \tag{6}$$

Square both sides of equations (4), (5) and (6) then introduce expectations to obtain

$$E(\sigma_{Xh})^2 = \theta_h(S_{Xh}^2 + S_{Vh}^2) + \theta_h^*(S_{Xh(2)}^2 + S_{Vh(2)}^2) = A_h$$

$$E(\sigma_{Yh})^2 = \theta_h(S_{Yh}^2 + S_{Uh}^2) + \theta_h^*(S_{Yh(2)}^2 + S_{Uh(2)}^2) = B_h$$

$$E(\sigma_{X1h})^2 = E(\sigma_{X1h}\sigma_{Xh}) = \theta_h' S_{Xh}^2 = C_h$$

$$E(\sigma_{X1h}\sigma_{Yh}) = \theta_h' \rho_{XYh} S_{Xh} S_{Yh} = D_h$$

$$E(\sigma_{Xh}\sigma_{Yh}) = \theta_h \rho_{XYh} S_{Xh} S_{Xh} + \theta_h^* \rho_{XYh(2)} S_{Xh} S_{Yh(2)} = E_h$$

, where $\theta_h' = \left(\frac{1}{n_h'} - \frac{1}{N_h}\right)$, $\theta_h = \left(\frac{1}{n_h} - \frac{1}{N_h}\right)$ and $\theta_h^* = \frac{W_{2h}(k_{2h}-1)}{n_h}$

$$E(\sigma_{Yh}) = E(\sigma_{Xh}) = E(\sigma_{X1h}) = 0$$

2.1 Bias and Mean Square Error of the Ratio Estimator

To obtain the expression of bias and mean square error of the ratio estimator substitute equations (4), (5) and (6) in T_R to obtain

$$T_R = \sum_{h=1}^{L} w_h(\bar{Y}_h + \sigma_{Yh})\left(1 + \frac{\sigma_{X1h}}{\bar{X}_h}\right)\left(1 + \frac{\sigma_{Xh}}{\bar{X}_h}\right)^{-1}$$

Simplify the right-hand side in terms of power series ignoring terms of order greater than 2 and subtract the population mean to obtain

$$(T_R - \bar{Y}) \approx \sum_{h=1}^{L} w_h\left[\sigma_{Yh} - R_h\sigma_{Xh} - \frac{\sigma_{Xh}\sigma_{Yh}}{\bar{X}_h} + \frac{R_h}{\bar{X}_h}\sigma_{Xh}^2 + R_h\sigma_{X1h} + \frac{\sigma_{X1h}\sigma_{Yh}}{\bar{X}_h} - \frac{R_h}{\bar{X}_h}\sigma_{Xh}\sigma_{X1h}\right] \tag{7}$$

Take expectations on both sides of equation (7) to obtain

$$E(T_R - \bar{Y}) \approx \sum_{h=1}^{L} W_h\left[E(\sigma_{Yh}) - R_h E(\sigma_{Xh}) - \frac{E(\sigma_{Xh}\sigma_{Yh})}{\bar{X}_h} + \frac{R_h}{\bar{X}_h}E(\sigma_{Xh}^2) + R_h E(\sigma_{X1h}) + \frac{E(\sigma_{X1h}\sigma_{Yh})}{\bar{X}_h} - \frac{R_h}{\bar{X}_h}E(\sigma_{Xh}\sigma_{X1h})\right]$$

The approximation of bias is given as

$$E(T_R - \bar{Y}) \approx \sum_{h=1}^{L} \frac{W_h}{\bar{X}_h}[R_h(A_h - C_h) + D_h - E_h]$$

To obtain the expression of mean square error, square equation (7) and ignore terms of order greater than 2

$$(T_R - \bar{Y})^2 \approx \sum_{h=1}^{L} w_h^2(\sigma_{Yh} - R_h\sigma_{Xh} + R_h\sigma_{X1h})^2$$

Simplify and take expectations on both sides to obtain the approximation of mean square error as

$$E(T_R - \bar{Y})^2 \approx \sum_{h=1}^{L} W_h^2[E(\sigma_{Yh}^2) + R_h^2 E(\sigma_{Xh}^2) - R_h^2 E(\sigma_{X1h}^2) + 2R_h E(\sigma_{Yh}\sigma_{X1h}) - 2R_h E(\sigma_{Yh}\sigma_{Xh})]$$

$$E(T_R - \bar{Y})^2 \approx \sum_{h=1}^{L} W_h^2[B_h + R_h^2(A_h - C_h) + 2R_h(D_h - E_h)]$$

This can be further simplified into
$MSE(T_R) \approx \sum_{h=1}^{L} W_h^2[\theta_h' V_{21} + \theta_h V_{22} + \theta_h^* V_{23}]$

, where $V_{21} = 2R_h\rho_{XYh}S_{Xh}S_{Yh} - R_h^2 S_{Xh}^2$

$V_{22} = S_{Yh}^2 + S_{Uh}^2 + R_h^2 S_{Xh}^2 + R_h^2 S_{Uh}^2 - 2R_h\rho_{XYh}S_{Xh}S_{Yh}$

$V_{23} = S_{Yh(2)}^2 + S_{Uh(2)}^2 + R_h^2 S_{Xh(2)}^2 + R_h^2 S_{Vh(2)}^2 - 2R_h\rho_{XYh(2)}S_{Xh(2)}S_{Yh(2)}$

2.2 Bias and Mean Square Error of the Exponential Ratio-ratio Type Estimator

To obtain the expression of bias and mean square error of the exponential ratio-ratio type estimator substitute equation (4), (5) and (6) in T_{ERR} to obtain

$$T_{ERR} = \sum_{h=1}^{L} w_h(\bar{Y}_h + \sigma_{Yh})\left(1 + \frac{\sigma_{X1h}}{\bar{X}_h}\right)\left(1 + \frac{\sigma_{Xh}}{\bar{X}_h}\right)^{-1} \exp\left[\frac{\sigma_{X1h} - \sigma_{Xh}}{2\bar{X}_h + \sigma_{X1h} + \sigma_{Xh}}\right]$$

Simplify while ignoring terms of order greater than 2 and subtract the population mean from both sides to obtain

$$(T_{ERR} - \bar{Y}) \approx \sum_{h=1}^{L} w_h\left[\sigma_{Yh} - \frac{3}{2}R_h\sigma_{Xh} - \frac{3}{2}\frac{\sigma_{Xh}\sigma_{Yh}}{\bar{X}_h} + \frac{15R_h}{8\bar{X}_h}\sigma_{Xh}^2 + \frac{3}{2}R_h\sigma_{X1h} + \frac{3}{2}\frac{\sigma_{X1h}\sigma_{Yh}}{\bar{X}_h} - \frac{9}{4}\frac{R_h}{\bar{X}_h}\sigma_{Xh}\sigma_{X1h} + \frac{3}{8}\frac{R_h}{\bar{X}_h}\sigma_{X1h}^2\right]$$
(8)

Take expectations on both sides of equation (8) to obtain the approximation of bias as

$$E(T_{ERR} - \bar{Y}) \approx \sum_{h=1}^{L} \frac{W_h}{2\bar{X}_h}\left[\frac{15}{4}R_h(A_h - C_h) + 3(D_h - E_h)\right]$$

To obtain the expression of mean square error, square both sides of equation (8) and ignore terms of order greater than 2

$$(T_{ERR} - \bar{Y})^2 \approx \sum_{h=1}^{L} w_h^2\left[\sigma_{Yh} - \frac{3}{2}R_h\sigma_{Xh} + \frac{3}{2}R_h\sigma_{X1h}\right]^2$$

Simplify and take expectations on both sides to obtain

$$E(T_{ERR} - \bar{Y})^2 \approx \sum_{h=1}^{L} W_h^2\left[E(\sigma_{Yh}^2) + \frac{9}{4}R_h^2E(\sigma_{Xh}^2) - \frac{9}{4}R_h^2E(\sigma_{X1h}^2) + 3R_hE(\sigma_{Yh}\sigma_{X1h}) - 3R_hE(\sigma_{Yh}\sigma_{Xh})\right]$$

$$E(T_{ERR} - \bar{Y})^2 \approx \sum_{h=1}^{L} W_h^2\left[B_h + \frac{9}{4}R_h^2(A_h - C_h) + 3R_h(D_h - E_h)\right]$$

This can be further simplified into

$$MSE(T_{ERR}) \approx \sum_{h=1}^{L} W_h^2[\theta_h' V_{31} + \theta_h V_{32} + \theta_h^* V_{33}]$$

, where $V_{31} = R_h\left(3\rho_{XYh}S_{Xh}S_{Yh} - \frac{9}{4}R_hS_{Xh}^2\right)$

$V_{32} = S_{Yh}^2 + S_{Uh}^2 + \frac{9}{4}R_h^2S_{Xh}^2 + \frac{9}{4}R_h^2S_{Vh}^2 - 3R_h\rho_{XYh}S_{Xh}S_{Yh}$

$V_{33} = S_{Yh(2)}^2 + S_{Uh(2)}^2 + \frac{9}{4}R_h^2S_{Xh(2)}^2 + \frac{9}{4}R_h^2S_{Vh(2)}^2 - 3R_h\rho_{XYh(2)}S_{Xh(2)}S_{Yh(2)}$

2.3 Bias and Mean Square Error of the Exponential Product-ratio Type Estimator

To obtain the expression of bias and mean square error of the exponential product- ratio type estimator substitute equation (4), (5) and (6) in T_{EPR} to obtain

$$T_{EPP} = \sum_{h=1}^{L} w_h(\bar{Y}_h + \sigma_{Yh})\left(1 + \frac{\sigma_{X1h}}{\bar{X}_h}\right)\left(1 + \frac{\sigma_{Xh}}{\bar{X}_h}\right)^{-1} \exp\left[\frac{\sigma_{Xh} - \sigma_{X1h}}{2\bar{X}_h + \sigma_{X1h} + \sigma_{Xh}}\right]$$

Simplify the right-hand side in terms of power series ignoring terms of order greater than two and subtract the population mean from both sides to obtain

$$(T_{EPP} - \bar{Y}) \approx \sum_{h=1}^{L} w_h\left[\sigma_{Yh} + \frac{1}{2}R_h\sigma_{X1h} + \frac{\sigma_{X1h}\sigma_{Yh}}{2\bar{X}_h} - \frac{1}{2}R_h\sigma_{Xh} - \frac{\sigma_{Xh}\sigma_{Yh}}{2\bar{X}_h} - \frac{R_h}{8\bar{X}_h}\sigma_{X1h}^2 + \frac{3R_h}{8\bar{X}_h}\sigma_{Xh}^2 - \frac{R_h}{4\bar{X}_h}\sigma_{Xh}\sigma_{X1h}\right]$$
(9)

Take expectations on both sides of equation (9) to obtain the approximation of bias as

$$E(T_{EPR} - \bar{Y}) \approx \sum_{h=1}^{L} \frac{W_h}{2\bar{X}_h}\left[\frac{3}{4}R_h(A_h - C_h) + (D_h - E_h)\right]$$

To obtain the expression of mean square error, square equation (9) and ignore terms of order greater than 2

$$(T_{EPR}-\bar{Y}) \approx \sum_{h=1}^{L} w_h^2 \left[\sigma_{Yh}-\frac{1}{2}R_h\sigma_{Xh}+\frac{1}{2}R_h\sigma_{X1h}\right]^2$$

Simplify and take expectations on both sides to obtain the approximation of mean square error as

$$E(T_{EPR}-\bar{Y})^2 \approx \sum_{h=1}^{L} W_h^2 \left[E(\sigma_{Yh}^2)+\frac{1}{4}R_h^2E(\sigma_{Xh}^2)-\frac{1}{4}R_h^2E(\sigma_{X1h}^2)+R_hE(\sigma_{Yh}\sigma_{X1h})-R_hE(\sigma_{Yh}\sigma_{Xh})\right]$$

$$E(T_{EPR}-\bar{Y})^2 \approx \sum_{h=1}^{L} W_h^2 \left[B_h+\frac{1}{4}R_h^2(A_h-C_h)+R_h(D_h-E_h)\right]$$

This can be further simplified into

$$MSE(T_{EPR}) \approx \sum_{h=1}^{L} W_h^2[\theta_h' V_{41}+\theta_h V_{42}+\theta_h^* V_{43}]$$

, where $V_{41}=R_h\left(\rho_{XYh}S_{Xh}S_{Yh}-\frac{1}{4}R_hS_{Xh}^2\right)$

$$V_{42}=S_{Yh}^2+S_{Uh}^2+\frac{1}{4}R_h^2S_{Xh}^2+\frac{1}{4}R_h^2S_{Vh}^2-R_h\rho_{Xyh}S_{Xh}S_{Yh}$$

$$V_{43}=S_{Yh(2)}^2+S_{Uh(2)}^2+\frac{1}{4}R_h^2S_{Xh(2)}^2+\frac{1}{4}R_h^2S_{Vh(2)}^2-R_h\rho_{XYh(2)}S_{Xh(2)}S_{Yh(2)}$$

3. Optimal Allocation in the Presence of Nonresponse and Measurement Error

Define the total cost of the survey in the h^{th} stratum as

$$C_h=c_h'n_h'+c_hn_h+c_{1h}n_{1h}+c_{2h}r_{2h},$$

, where

c_h' = Cost of measuring a unit in the first phase stratum sample of size n_h'

c_h = Cost of measuring a unit in the second phase stratum sample of size n_h

c_{1h} = unit cost of processing respondent data in the first attempt of size n_{1h}

c_{2h} = unit cost of processing data in the subsample of size r_{2h} obtained from the nonresponding units.

The values of n_{1h} and r_{2h} are unknown until the first attempt is done. Let $w_{1h}=\frac{n_{1h}}{n_h}$, $w_{2h}=\frac{n_{2h}}{n_h}$ and $r_{2h}=\frac{n_{2h}}{k_{2h}}$. Therefore the expected cost function is defined as

$$C_h^*=c_h'n_h'+n_h\left(c_h+c_{1h}W_{1h}+c_{2h}\frac{W_{2h}}{k_{2h}}\right)$$

The expected total cost function is given as

$$C^*=C_h^*$$

To obtain the optimal allocation for the ratio estimator of the population mean define the Lagrangian function for optimization as

$$\phi=\left\{\left(\frac{1}{n_h'}-\frac{1}{N_h}\right)V_{21}+\left(\frac{1}{n_h}-\frac{1}{N_h}\right)V_{22}+\left(\frac{W_{2h}(k_{2h}-1)}{n_h}\right)V_{23}\right\}+\lambda\left[c_h'n_h'+n_h\left(c_h+c_{1h}n_hW_{1h}+c_{2h}\frac{W_{2h}}{k_{2h}}\right)-C_h^*\right],$$

, where λ is a Lagrange's multiplier. The optimum values of n_h', n_h and k_{2h} are obtained by differentiating the Lagrangian function partially with respect to n_h', n_h and k_{2h} and equating the derivatives to zero as follows

$$\frac{\partial\phi}{\partial n_h'}=-\frac{1}{n_h'^2}V_{21}+\lambda c_h'=0$$

$$n_h'=\sqrt{\frac{V_{21}}{\lambda c_h'}}$$

Next differentiate the Lagrangian function with respect to k_{2h}

$$\frac{\partial\phi}{\partial k_{2h}}=\frac{W_{2h}}{n_h}V_{23}-\lambda\frac{c_{2h}W_{2h}n_h}{k_{2h}^2}=0$$

$$k_{2h}=n_h\sqrt{\lambda\frac{c_{2h}}{V_{23}}}$$

$$\frac{k_{2h}}{n_h}=\sqrt{\lambda\frac{c_{2h}}{V_{23}}} \qquad (10)$$

$$\frac{n_h}{k_{2h}}=\sqrt{\frac{V_{23}}{\lambda c_{2h}}} \qquad (11)$$

The optimum value of n_h is obtained by substituting equations (10) and (11) in the Lagrangian function and differentiating it partially with respect to n_h.

$$\frac{\partial\phi}{\partial n_h} = -\frac{1}{n_h^2}V_{22} + \frac{W_{2h}}{n_h^2}V_{23} + \lambda(c_h + c_{1h}W_{1h}) = 0$$

$$n_h = \sqrt{\frac{V_{22}-V_{23}W_{2h}}{\lambda(c_h+c_{1h}W_{1h})}}$$

Substitute the above equation in the value of k_{2h} to obtain its optimum value as

$$k_{2h}^* = \sqrt{\frac{c_{2h}(V_{22}-V_{23}W_{2h})}{V_{23}(c_h+c_{1h}W_{1h})}}$$

The Lagrange's multiplier is obtained by substituting the values of n'_h, n_h and k_{2h}^* in the expected cost function and is defined as

$$\sqrt{\lambda} = \frac{1}{C^*}\sum_{h=1}^{L}\left[c'_h\sqrt{\frac{V_{21}}{c'_h}} + \left(c_h + c_{1h}W_{1h} + \frac{c_{2h}W_{2h}}{k_{2h}^*}\right)\left(\sqrt{\frac{(V_{22}-V_{23}W_{2h})}{(c_h+c_{1h}W_{1h})}}\right)\right]$$

$$n'_{h_{opt}} = C^* \times \left(\sqrt{\frac{V_{21}}{c'_h}}\right) \times \sum_{h=1}^{L}\left[c'_h\sqrt{\frac{V_{21}}{c'_h}} + \left(c_h + c_{1h}W_{1h} + \frac{c_{2h}W_{2h}}{k_{2h}^*}\right)\left(\sqrt{\frac{(V_{22}-V_{23}W_{2h})}{(c_h+c_{1h}W_{1h})}}\right)\right]^{-1}$$

$$n_{h_{opt}} = C^* \times \left(\sqrt{\frac{V_{22}-V_{23}W_{2h}}{(c_h+c_{1h}W_{1h})}}\right) \times \sum_{h=1}^{L}\left[c'_h\sqrt{\frac{V_{21}}{c'_h}} + \left(c_h + c_{1h}W_{1h} + \frac{c_{2h}W_{2h}}{k_{2h}^*}\right)\left(\sqrt{\frac{(V_{22}-V_{23}W_{2h})}{(c_h+c_{1h}W_{1h})}}\right)\right]^{-1}$$

The minimum mean square error of the ratio estimator is given as

$$MSE(T_R)_{min} = \frac{1}{C^*}\sum_{h=1}^{L}W_h^2\left\{\left[V_{21}\sqrt{\frac{c'_h}{V_{21}}} + V_{23}W_{2h}\sqrt{\frac{c_{2h}}{V_{23}}} + (V_{22}-V_{23}W_{2h})\sqrt{\frac{c_h+c_{1h}W_{1h}}{V_{22}-V_{23}W_{2h}}}\right]\right.$$
$$\left.\times\left[c'_h\sqrt{\frac{V_{21}}{c'_h}} + \left(c_h + c_{1h}W_{1h} + \frac{c_{2h}W_{2h}}{k_{2h}^*}\right)\left(\sqrt{\frac{(V_{22}-V_{23}W_{2h})}{(c_h+c_{1h}W_{1h})}}\right)\right]\right\}$$

To obtain the optimal allocation for the exponential ratio-ratio type estimator define the Lagrangian function for optimization as

$$\phi = \left\{\left(\frac{1}{n'_h}-\frac{1}{N_h}\right)V_{31} + \left(\frac{1}{n_h}-\frac{1}{N_h}\right)V_{32} + \left(\frac{W_{2h}(k_{2h}-1)}{n_h}\right)V_{33}\right\} + \lambda\left[c'_h n'_h + n_h\left(c_h + c_{1h}n_hW_{1h} + c_{2h}\frac{W_{2h}}{k_{2h}}\right) - C_h^*\right]$$

, where λ is a Lagrange's multiplier. To obtain the optimum values of n'_h, n_h and k_{2h} differentiate the Lagrangian function partially with respect to n'_h, n_h and k_{2h} then equate the derivatives to zero

$$\frac{\partial\phi}{\partial n'_h} = -\frac{1}{n'^2_h}V_{31} + \lambda c'_h = 0$$

$$n'_h = \sqrt{\frac{V_{31}}{\lambda c'_h}}$$

Next differentiate the Lagrangian function with respect to k_{2h}

$$\frac{\partial\phi}{\partial k_{2h}} = \frac{W_{2h}}{n_h}V_{33} - \lambda\frac{c_{2h}W_{2h}n_h}{k_{2h}^2} = 0$$

$$k_{2h} = n_h\sqrt{\lambda\frac{c_{2h}}{V_{33}}}$$

$$\frac{k_{2h}}{n_h} = \sqrt{\lambda\frac{c_{2h}}{V_{33}}} \qquad (12)$$

$$\frac{n_h}{k_{2h}} = \sqrt{\frac{V_{33}}{\lambda c_{2h}}} \qquad (13)$$

The optimum value of n_h is obtained by substituting equations (12) and (13) in the Lagrangian function and differentiating it partially with respect to n_h.

$$\frac{\partial\phi}{\partial n_h} = -\frac{1}{n_h^2}V_{32} + \frac{W_{2h}}{n_h^2}V_{33} + \lambda(c_h + c_{1h}W_{1h}) = 0$$

$n_h = \sqrt{\frac{V_{32}-V_{33}W_{2h}}{\lambda(c_h+c_{1h}W_{1h})}}$

Substitute the above equation in the value of k_{2h} to obtain its optimum value as

$k_{2h}^* = \sqrt{\frac{c_{2h}(V_{32}-V_{33}W_{2h})}{V_{33}(c_h+c_{1h}W_{1h})}}$

The Lagrange's multiplier is obtained by substituting the values of n_h', n_h and k_{2h}^* in the expected total cost function

$$\sqrt{\lambda} = \frac{1}{C^*}\sum_{h=1}^{L}\left[c_h'\sqrt{\frac{V_{31}}{c_h'}} + \left(c_h + c_{1h}W_{1h} + \frac{c_{2h}W_{2h}}{k_{2h}^*}\right)\left(\sqrt{\frac{(V_{32}-V_{33}W_{2h})}{(c_h+c_{1h}W_{1h})}}\right)\right]$$

$$n'_{h_{opt}} = C^* \times \left(\sqrt{\frac{V_{31}}{c_h'}}\right) \times \sum_{h=1}^{L}\left[c_h'\sqrt{\frac{V_{31}}{c_h'}} + \left(c_h + c_{1h}W_{1h} + \frac{c_{2h}W_{2h}}{k_{2h}^*}\right)\left(\sqrt{\frac{(V_{32}-V_{33}W_{2h})}{(c_h+c_{1h}W_{1h})}}\right)\right]^{-1}$$

$$n_{h_{opt}} = C^* \times \left(\sqrt{\frac{V_{32}-V_{33}W_{2h}}{(c_h+c_{1h}W_{1h})}}\right) \times \sum_{h=1}^{L}\left[c_h'\sqrt{\frac{V_{31}}{c_h'}} + \left(c_h + c_{1h}W_{1h} + \frac{c_{2h}W_{2h}}{k_{2h}^*}\right)\left(\sqrt{\frac{(V_{32}-V_{33}W_{2h})}{(c_h+c_{1h}W_{1h})}}\right)\right]^{-1}$$

The minimum mean square error of the exponential ratio-ratio type estimator is given as

$$MSE(T_{ERR})_{min} = \frac{1}{C^*}\sum_{h=1}^{L}W_h^2\left\{\left[V_{31}\sqrt{\frac{c_h'}{V_{31}}} + V_{33}W_{2h}\sqrt{\frac{c_{2h}}{V_{33}}} + (V_{32}-V_{33}W_{2h})\sqrt{\frac{c_h+c_{1h}W_{1h}}{V_{32}-V_{33}W_{2h}}}\right] \times \left[c_h'\sqrt{\frac{V_{31}}{c_h'}} + \left(c_h + c_{1h}W_{1h} + \frac{c_{2h}W_{2h}}{k_{2h}^*}\right)\left(\sqrt{\frac{(V_{32}-V_{33}W_{2h})}{(c_h+c_{1h}W_{1h})}}\right)\right]\right\}$$

To obtain the optimal allocation for the exponential product-ratio type estimator define the Lagrangian function for optimization as

$$\phi = \left\{\left(\frac{1}{n_h'}-\frac{1}{N_h}\right)V_{41} + \left(\frac{1}{n_h}-\frac{1}{N_h}\right)V_{42} + \left(\frac{W_{2h}(k_{2h}-1)}{n_h}\right)V_{43}\right\} + \lambda\left[c_h'n_h' + n_h\left(c_h + c_{1h}n_hW_{1h} + c_{2h}\frac{W_{2h}}{k_{2h}}\right) - C_h^*\right]$$

, where λ is a Lagrange's multiplier. To obtain the optimum values of n_h', n_h and k_{2h} differentiate the Lagrangian function partially with respect to n_h', n_h and k_{2h} then equate the derivatives to zero

$$\frac{\partial\phi}{\partial n_h'} = -\frac{1}{n_h'^2}V_{41} + \lambda c_h' = 0$$

$$n_h' = \sqrt{\frac{V_{41}}{\lambda c_h'}}$$

Next, differentiate the Lagrangian function with respect to k_{2h}

$$\frac{\partial\phi}{\partial k_{2h}} = \frac{W_{2h}}{n_h}V_{43} - \lambda\frac{c_{2h}W_{2h}n_h}{k_{2h}^2} = 0$$

$$k_{2h} = n_h\sqrt{\lambda\frac{c_{2h}}{V_{43}}}$$

$$\frac{k_{2h}}{n_h} = \sqrt{\lambda\frac{c_{2h}}{V_{43}}} \qquad (14)$$

$$\frac{n_h}{k_{2h}} = \sqrt{\frac{V_{43}}{\lambda c_{2h}}} \qquad (15)$$

The optimum value of n_h is obtained by substituting equations (14) and (15) in the Lagrangian function and differentiating it partially with respect to n_h.

$$\frac{\partial\phi}{\partial n_h} = -\frac{1}{n_h^2}V_{42} + \frac{W_{2h}}{n_h^2}V_{43} + \lambda(c_h + c_{1h}W_{1h}) = 0$$

$$n_h = \sqrt{\frac{V_{42}-V_{43}W_{2h}}{\lambda(c_h+c_{1h}W_{1h})}}$$

Substitute the above equation in the value of k_{2h} to obtain its optimum value as

$$k_{2h}^{*}=\sqrt{\frac{c_{2h}(V_{42}-V_{43}W_{2h})}{V_{43}(c_h+c_{1h}W_{1h})}}$$

The Lagrange's multiplier is obtained by substituting the values of n_h', n_h and k_{2h}^{*} in the expected cost function

$$\sqrt{\lambda}=\frac{1}{C^{*}}\sum_{h=1}^{L}\left[c_h'\sqrt{\frac{V_{41}}{c_h'}}+\left(c_h+c_{1h}W_{1h}+\frac{c_{2h}W_{2h}}{k_{2h}^{*}}\right)\left(\sqrt{\frac{(V_{42}-V_{43}W_{2h})}{(c_h+c_{1h}W_{1h})}}\right)\right]$$

$$n_{h_{opt}}'=C^{*}\times\left(\sqrt{\frac{V_{41}}{c_h'}}\right)\times\sum_{h=1}^{L}\left[c_h'\sqrt{\frac{V_{41}}{c_h'}}+\left(c_h+c_{1h}W_{1h}+\frac{c_{2h}W_{2h}}{k_{2h}^{*}}\right)\left(\sqrt{\frac{(V_{42}-V_{43}W_{2h})}{(c_h+c_{1h}W_{1h})}}\right)\right]^{-1}$$

$$n_{h_{opt}}=C^{*}\times\left(\sqrt{\frac{V_{42}-V_{43}W_{2h}}{(c_h+c_{1h}W_{1h})}}\right)\times\sum_{h=1}^{L}\left[c_h'\sqrt{\frac{V_{41}}{c_h'}}+\left(c_h+c_{1h}W_{1h}+\frac{c_{2h}W_{2h}}{k_{2h}^{*}}\right)\left(\sqrt{\frac{(V_{42}-V_{43}W_{2h})}{(c_h+c_{1h}W_{1h})}}\right)\right]^{-1}$$

The minimum mean square error of the exponential product-ratio type estimator is given as

$$MSE(T_{EPR})_{min}=\frac{1}{C^{*}}\sum_{h=1}^{L}W_h^2\left\{\left[V_{41}\sqrt{\frac{c_h'}{V_{41}}}+V_{43}W_{2h}\sqrt{\frac{c_{2h}}{V_{43}}}+(V_{42}-V_{43}W_{2h})\sqrt{\frac{c_h+c_{1h}W_{1h}}{V_{42}-V_{43}W_{2h}}}\right]\right.$$
$$\left.\times\left[c_h'\sqrt{\frac{V_{41}}{c_h'}}+\left(c_h+c_{1h}W_{1h}+\frac{c_{2h}W_{2h}}{k_{2h}^{*}}\right)\left(\sqrt{\frac{(V_{42}-V_{43}W_{2h})}{(c_h+c_{1h}W_{1h})}}\right)\right]\right\}$$

4. Conclusion

The present study has proposed a ratio, exponential ratio-ratio type and exponential product-ratio type estimators of the population mean in the presence of nonresponse and measurement errors under double sampling for stratification. The expression of mean square errors and biases of the estimators have been derived. The cost of the survey has been studied theoretically. The optimum values of the sample sizes and the inverse sampling rate have been derived. It is noted that the optimum value of the sample size depends on the value of the Lagrange's multiplier. The mean square error of the estimators decreases with the increase in the cost of the survey. A high cost of the survey results in the selection of a large sample size.

5. References

1. Javid S, Sat G, Shakeel A. A generalized class of estimators under two-phase stratified sampling for non-response. Communications in Statistics-Theory and Methods, 2018. DOI:10.1080/03610926.2018.1481969.
2. Searls DT. The utilization of known coefficient of variation in the estimation procedure. Journal of American Statistical Association. 1964; 308:1225-1226.
3. Zahid E, Shabbir J. Estimation of population mean in the presence of measurement error and non-response under stratified random sampling. PLoS ONE. 2018; 2:e0191572. https://doi.org/10.1371/journal.pone.0191572.
4. Onyango OR, Ouma OC. Application of auxiliary variables in two-step semi-parametric multiple imputation procedure in estimation of population mean. International Journal of Science and Research (IJSR). 2017; 6: 77-83. DOI: 10.21275/ART20176978.
5. Cochran W. Sampling technique, Third Edition, John Wiley and Sons, New York, 1977.
6. Shalabh. Ratio method of estimation in the presence of measurement errors. Journal of Indian Society of Agricultural Statistics. 1997; 50:150-155.
7. Shukla D, Pathak S, Thakur N. An estimator for mean estimation in presence of measurement error. Research and Reviews: A Journal of Statistics. 2012; 1:1-8.54.
8. Singh H, Karpe N. Estimation of population variance using auxiliary information in the presence of measurement error. Statistics in Transition-New Series. 2008; 9:443-470.
9. Hansen M, Hurwitz W. The problem of non-response in sample surveys. Journal of American Statistical Association. 1946; 41:517-529.