Estimation of population mean in the presence of measurement error and non-response under stratified random sampling

Erum Zahid*, Javid Shabbir

Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan

* erumzahid22@gmail.com

Abstract

In the present paper we propose an improved class of estimators in the presence of measurement error and non-response under stratified random sampling for estimating the finite population mean. The theoretical and numerical studies reveal that the proposed class of estimators performs better than other existing estimators.

Introduction

In survey sampling, usually it is assumed that all the observations are correctly measured on the characteristics under study. But in practice this assumption is not met for a variety of reasons, such as non-response may occurs due to refusal of respondents to give the information or not at home or lack of interest or due to some ethical reasons. Usually measurement error and non-response are studied separately using the known auxiliary or additional information. In reality, both measurement error and non-response occur simultaneously in survey sampling. Mostly the information is not obtained from all the units during surveys, so non-response is a common problem which may creeps during a sample survey. In sampling theory the estimation of population mean of a variable of interest in the presence of non-response, when the auxiliary information available is widely debated. [1–5] and [6] discussed the problems of non-response in detail. To estimate the population mean, the researchers dealt with the problem of measurement error. For more details, see [7–11], etc. Recently few researchers studied the problem of measurement error and non-response together like [12–14] and [15]. [16] and [17] studied the improved estimation of population mean in simple and stratified random sampling.

In practice, the researchers who have studied measurement error have ignored the presence of non-response. But very few of them studied both under simple random sampling. In this paper, we have proposed a class of estimators for estimating the population mean in the presence of measurement error and non-response simultaneously under stratified random sampling. The efficiency of the suggested class of estimators over the existing estimators is shown through simulation study and real data sets.

Consider a finite population of \( N \) identifiable units which are partitioned into \( L \) homogeneous subgroups called strata, such that the \( h^{th} \) strata consist of \( N_h \) units, where \( h = 1, 2, \ldots, L \).
and \( \sum_{h=1}^{k} N_h = N \). It is assumed that \( N \) consists of two mutually exclusive groups called response and non-response groups. Let \( N_{1h} \) and \( N_{2h} \) are the responding and non-responding unit in the \( h^{th} \) stratum respectively. We select a sample of size \( n_h \) from \( \mathcal{N}_h \) by using simple random sampling without replacement (SRSWOR) and assume that \( n_{1h} \) units respond and \( n_{2h} \) units do not respond. We select a sub-sample of size \( k_{ih} \left( k_{ih} = \frac{n_{ih}}{n_{ih}} \right) \) from \( n_{2h} \) non-responding units in the \( h^{th} \) stratum.

Let \((x_{ih}^*, y_{ih}^*)\) be the observed values and \((X_{ih}, Y_{ih})\) be the actual values on the variables \((X, Y)\) of the \( i^{th} \) \((i = 1, 2, \ldots, n)\) sample units in the \( h^{th} \) stratum. Then the measurement errors be \( V_{ih} = x_{ih}^* - X_{ih} \) and \( U_{ih} = y_{ih}^* - Y_{ih} \). Let \( S_{hV}^2 \) and \( S_{hX}^2 \) be the population variances for the responding units, and \( S_{hY(2)}^2 \) and \( S_{hX(2)}^2 \) be the population variances for non-responding units. Let \( S_{hV}^2 \) and \( S_{hU}^2 \) be the population variances associated with the measurement error for responding units. Let \( S_{hY(2)}^2 \) and \( S_{hX(2)}^2 \) be the population variances associated with measurement error for the non responding part of the population. Further let \( C_{hY} \) and \( C_{hX} \) be the coefficient of variations for the respondents and \( C_{hY(2)} \) and \( C_{hX(2)} \) be the coefficient of variations for the non-responding units respectively. Let \( \rho_{hYX} \) and \( \rho_{hYX(2)} \) be the population correlation coefficients between their respective subscripts respectively for responding and non-responding units, respectively.

In this paper an improved class of estimators for estimating the population mean in the presence of measurement error and non-response is proposed under stratified random sampling. Expressions for the bias and mean square error (MSE) of the class of estimators are obtained upto first order of approximation, when both the study and the auxiliary variables suffer with a problem of non-response and measurement errors.

The present paper is organized as: Section 2 gives some existing estimators of the finite population mean. In Section 3, an improved class of estimators is suggested for estimating the finite population mean by incorporating both measurement error and non-response information simultaneously. Efficiency comparison is presented in section 4. Numerical results and simulation study are presented in Section 5. Conclusion is given in Section 6.

**Existing estimators in literature**

In this section we consider the following existing estimators.

**Hansen and Hurwitz (1946) estimator**

In stratified random sampling, the Hansen and Hurwitz (1946) estimator for population mean \( \mathcal{Y} \), is given by

\[
\bar{Y}_{S(HH)} = \sum_{h=1}^{k} p_h \bar{Y}^*_h, \tag{1}
\]

where \( \bar{Y}^*_h = \frac{n_{ih}}{n_h} \bar{Y}_{1h} + \frac{n_{ih}}{n_h} \bar{Y}_{2h} \) and \( p_h = \frac{n_{ih}}{n_h} \).

Here \( \bar{Y}_{1h} \) and \( \bar{Y}_{2h} \) are the sample means based on \( n_{1h} \) of responding and \( n_{2h} \) units of subsamples from \( n_{2h} \) non-responding groups, respectively.

The variance of \( \bar{Y}_{S(HH)} \), is given by

\[
\text{Var}(\bar{Y}_{S(HH)}) = \sum_{h=1}^{k} p_h^2 A_h, \tag{2}
\]

where

\[
A_h = \lambda_{2h} \left( S_{hY}^2 + S_{hU}^2 \right) + \theta_h \left( S_{hY(2)}^2 + S_{hX(2)}^2 \right), \quad \theta_h = \frac{n_{ih}}{n_h}, \quad p_h = \frac{n_{ih}}{n_h}, \quad \lambda_{2h} = \left( n_h^{-1} - N_h^{-1} \right).
\]
**Ratio estimator**

The usual ratio estimator under stratified random sampling, is given by

\[
\bar{y}_{S(R)} = \sum_{h=1}^{t} P_h \frac{y_h}{x_h} \bar{x}_h,
\]

where \(\bar{x}_h\) is known and \(y_h\) is given in Eq (13). The bias and mean square error of \(\bar{y}_{S(R)}\), are given by

\[
B(\bar{y}_{S(R)}) \approx \sum_{h=1}^{t} P_h \frac{1}{x_h} [R_h B_h - C_h]
\]

and

\[
MSE(\bar{y}_{S(R)}) \approx \sum_{h=1}^{t} P_h^2 [A_h + R_h^2 B_h - 2R_h C_h],
\]

where \(R_h = \sum_{h'=1}^{t} B_{h'} = \lambda_{1h} (S_{hX}^2 + S_{hY}^2) + \theta_{h} (S_{hXY(2)}^2 + S_{hXY(2)}^2), C_h = \lambda_{2h} \rho_{hX} S_{hY} S_{hX} + \theta_{h} \rho_{hXY(2)} S_{hY(2)} S_{hX(2)}\).

**Difference estimator**

The usual difference estimator under stratified random sampling, is given by

\[
\bar{y}_{S(D)} = \sum_{h=1}^{t} P_h [y_h + d_h (x_h - \bar{x}_h)],
\]

where \(\bar{x}_h = \frac{N_h \bar{x}_h}{n_h} - \bar{x}_h\) and \(d_h\) is the constant.

The minimum variance of \(\bar{y}_{S(D)}\) is given by

\[
\text{Var}(\bar{y}_{S(D)})_{\text{min}} = \sum_{h=1}^{t} P_h^2 \left[ A_h - \frac{C_h^2}{B_h} \right],
\]

The optimum value of \(d_h\) is \(d_h^{(opt)} = - \frac{C_h}{B_h} t_h\) where \(t_h = \frac{m_h}{N_h - n_h}\).

**Azeem and Hanif (2017) estimator**

Azeem and Hanif (2016) estimator under stratified random sampling, is given by

\[
\bar{y}_{S(AH)} = \sum_{h=1}^{t} P_h \bar{y}_h \left( \frac{x_h}{\bar{x}_h} \right) \left( \frac{\bar{x}_h}{x_h} \right).
\]

The bias and MSE of \(\bar{y}_{S(AH)}\), are given by

\[
B(\bar{y}_{S(AH)}) \approx \sum_{h=1}^{t} P_h \frac{1}{\bar{x}_h} [R_h B_h - q_h C_h]
\]
and

\[ \text{MSE}(\hat{y}_{S(P)}) \cong \sum_{h=1}^{L} P_h [A_h + q_h^2 R_h^2 B_h - 2q_h R_h C_h], \]

where \( q_h = \frac{N_h - n_h}{N_h - m_h} \).

**The suggested estimator**

We propose an improved general class of estimators for estimating the population mean, dealing with the problem of measurement error and non-response simultaneously in stratified random sampling. Measurement error and non-response is present on both, the study and the auxiliary variables. The suggested estimator is given by

\[
\hat{y}_{S(P)} = \sum_{h=1}^{L} P_h \left[ \left( m_{1h} \bar{y}_h + m_{2h} (\bar{x}_h - \bar{y}_h) \right) \frac{X_h}{\bar{x}_h} \right] \exp \left( 1 - \alpha_h \right) \left( \frac{\bar{y}_h - \bar{x}_h}{\bar{x}_h + \bar{x}_h^*} \right),
\]

where, \( m_{1h} \) and \( m_{2h} \) are constants whose values are to be determined and \( \alpha_h \) is the scalar, chosen arbitrarily. For obtaining the bias and mean square error, we assume that

\[
\delta_{hY} = \sum_{i=1}^{n_h} (Y_{hi}^* - \bar{y}_h), \quad \delta_{hU} = \sum_{i=1}^{n_h} U_{hi}^*, \quad \delta_{hX} = \sum_{i=1}^{n_h} (X_{hi}^* - \bar{x}_h), \quad \delta_{hV} = \sum_{i=1}^{n_h} V_{hi}^*.
\]

Adding \( \delta_{hY} \) and \( \delta_{hU} \), we get \( \delta_{hY} + \delta_{hU} = \sum_{i=1}^{n_h} (Y_{hi}^* - \bar{y}_h) + \sum_{i=1}^{n_h} U_{hi}^* \).

Dividing both sides by \( n_h \), and then simplifying, we get

\[
\hat{y}_h = \bar{y}_h + \frac{1}{n_h} (\delta_{hY} + \delta_{hU}).
\]

Similarly, we can get

\[
\bar{x}_h = \bar{x}_h + \frac{1}{n_h} (\delta_{hX} + \delta_{hV}).
\]

Further

\[
E \left( \frac{\hat{y}_h - \bar{y}_h}{n_h} \right)^2 = \lambda_{hY} (S_{hY} + S_{hU}) + \theta_{hY} (S_{hY(2)} + S_{hU(2)}) = A_h,
\]

\[
E \left( \frac{\hat{y}_h - \bar{x}_h}{n_h} \right)^2 = \lambda_{hX} (S_{hX} + S_{hY}) + \theta_{hX} (S_{hX(2)} + S_{hY(2)}) = B_h,
\]

\[
E \left( \frac{\hat{x}_h - \bar{x}_h}{n_h} \right) = \lambda_{hX} \rho_{hXY} S_{hY} S_{hX} + \theta_{hX} S_{hX(2)} S_{hY(2)} = C_h.
\]

On simplifying, we get

\[
\hat{y}_{S(P)} = \sum_{h=1}^{L} P_h \left[ \left( m_{1h} (\bar{y}_h + W_{hY}) + m_{2h} (t_h W_{hX}) \right) \left( 1 + \frac{\alpha_h W_{hX}}{\bar{x}_h} \right) \right] \left( \frac{t_h W_{hX}}{2\bar{x}_h^2} + \frac{t_h^2 W_{hX}^2}{4\bar{x}_h^3} \right),
\]

where \( W_{hY} = \frac{\delta_{hY} + \delta_{hU}}{n_h} \) and \( W_{hX} = \frac{\delta_{hX} + \delta_{hV}}{n_h} \).
Further simplifying, and ignoring error terms greater than two, we have

$$\bar{\bar{y}}_S(p) - \overline{\bar{y}} \cong \sum_{h=1}^{L} P_h \left[ (m_{1h} - 1) \bar{Y}_h + m_{1h} \left( W_{hY} + e_i h R_h W_{hX} + \frac{e_i h R_h t_i h W_{hX} W_{hY}}{X_h} \right) + f_i h R_h t_i h W_{iX} X_h \right] + m_{2h} \left( t_i h W_{hX} + \frac{e_i h t_i h W_{hX}^2}{X_h} \right),$$

(15)

where, \( e_h = \frac{1 + 2 \sigma_h + \beta_h}{2} \) and \( f_h = \frac{\sigma_h^2 + \beta_h^2}{2} \).

Using Eq (15), the bias of \( \bar{\bar{y}}_S(p) \) to first order of approximation, is given by

$$B(\bar{\bar{y}}_S(p)) \cong \sum_{h=1}^{L} P_h (m_{1h} - 1) \bar{Y}_h + m_{1h} \left( e_i h t_i h R_h C_h + \frac{f_i h R_h t_i h B_h}{X_h} \right) + m_{2h} \left( e_i h t_i h R_h B_h \right).$$

(16)

Squaring both sides of Eq (15), and keeping the terms up to power two in errors, and then taking expectations, the mean square error of \( \bar{\bar{y}}_S(p) \) is given by

$$\text{MSE}(\bar{\bar{y}}_S(p)) \cong \sum_{h=1}^{L} P_h \left[ \bar{Y}_h^2 + 2 m_{1h} m_{2h} (t_i h C_h + 2 e_i h t_i h R_h B_h) \right. \nonumber$$

$$\left. + m_{1h}^2 (Y_h^2 + A_{h1} + c_i h t_i h R_h B_h + 4 e_i h t_i h R_h C_h + 2 f_i h t_i h R_h B_h) + m_{2h}^2 t_i h B_h \right]$$

$$- 2 m_{1h} (Y_h^2 + e_i h t_i h R_h C_h + f_i h t_i h R_h B_h) - 2 m_{2h} e_i h t_i h R_h B_h].$$

The above equation can be written as

$$\text{MSE}(\bar{\bar{y}}_S(p)) \cong \sum_{h=1}^{L} P_h \left[ \bar{Y}_h^2 + m_{1h} A_{h1} + m_{2h}^2 B_{h1} + 2 m_{1h} m_{2h} C_{h1} - 2 m_{1h} D_{h1} - 2 m_{2h} E_{h1} \right],$$

(17)

where,

$$A_{h1} = Y_h^2 + A_{h1} + c_i h t_i h R_h B_h + 4 e_i h t_i h R_h C_h + 2 f_i h t_i h R_h B_h,$$

$$B_{h1} = t_i h^2 B_h,$$

$$C_{h1} = t_i h C_h + 2 e_i h t_i h R_h B_h,$$

$$D_{h1} = Y_h^2 + e_i h t_i h R_h C_h + f_i h t_i h R_h B_h,$$

$$E_{h1} = e_i h t_i h R_h B_h.$$

For finding the optimal values of \( m_{1h} \) and \( m_{2h} \), we differentiate Eq (17) with respect to \( m_{1h} \) and \( m_{2h} \), respectively. The optimal values are given by \( m_{1h(opt)} = \frac{\frac{a_{h1}}{c_{h1}} + \frac{d_{h1}}{b_{h1}}}{a_{h1} + \beta_{h1}} \) and

$$m_{2h(opt)} = \frac{d_{h1} - c_{h1} f_{h1}}{c_{h1} - a_{h1} f_{h1}}.$$

Substituting these optimum values in Eq (17), we get the minimum mean square error of \( \bar{\bar{y}}_S(p) \), as

$$\text{MSE}(\bar{\bar{y}}_S(p))_{\text{min}} \cong \sum_{h=1}^{L} P_h \left[ \bar{Y}_h^2 - \frac{A_{h1} E_{h1} + B_{h1} D_{h1} - 2 C_{h1} D_{h1} E_{h1}}{(A_{h1} B_{h1} - C_{h1}^2)} \right].$$

(18)
Efficiency comparison

The efficiency comparison of \( \bar{y}_{S(HH)} \), \( \bar{y}_{S(R)} \), \( \bar{y}_{S(D)} \) and \( \bar{y}_{S(AH)} \) with respect to \( \bar{y}_{S(P)} \) are given by,

1. From Eqs (2) and (18), \( \text{MSE}(\bar{y}_{S(P)}) \text{ min} < \text{Var}(\bar{y}_{S(HH)}) \) if
   \[
   \sum_{h=1}^{L} P_h \bar{y}_h^2 - \sum_{h=1}^{L} P_h \left[ \frac{a_{11}^2 e_{11}^2 + b_{11}^2 d_{11}^2 - 2c_{11}^1 d_{11} b_{11}}{(a_{11}^2 b_{11}^2 - c_{11}^2)} \right] - \sum_{h=1}^{L} P_h^2 A_h < 0
   \]

2. From Eqs (5) and (18), \( \text{MSE}(\bar{y}_{S(P)}) \text{ min} < \text{MSE}(\bar{y}_{S(R)}) \) if
   \[
   \sum_{h=1}^{L} P_h \bar{y}_h^2 - \sum_{h=1}^{L} P_h \left[ \frac{a_{11}^2 e_{11}^2 + b_{11}^2 d_{11}^2 - 2c_{11}^1 d_{11} b_{11}}{(a_{11}^2 b_{11}^2 - c_{11}^2)} \right] - \sum_{h=1}^{L} P_h^2 [A_h + R_h^2 B_h - 2R_h C_h] < 0
   \]

3. From Eqs (7) and (18), \( \text{MSE}(\bar{y}_{S(P)}) \text{ min} < \text{Var}(\bar{y}_{S(D)}) \text{ min} \) if
   \[
   \sum_{h=1}^{L} P_h \bar{y}_h^2 - \sum_{h=1}^{L} P_h \left[ \frac{a_{11}^2 e_{11}^2 + b_{11}^2 d_{11}^2 - 2c_{11}^1 d_{11} b_{11}}{(a_{11}^2 b_{11}^2 - c_{11}^2)} \right] - \sum_{h=1}^{L} P_h^2 [A_h - \frac{C_h^2}{P_h}] < 0
   \]

4. From Eqs (10) and (18), \( \text{MSE}(\bar{y}_{S(P)}) \text{ min} < \text{MSE}(\bar{y}_{S(AH)}) \) if
   \[
   \sum_{h=1}^{L} P_h \bar{y}_h^2 - \sum_{h=1}^{L} P_h \left[ \frac{a_{11}^2 e_{11}^2 + b_{11}^2 d_{11}^2 - 2c_{11}^1 d_{11} b_{11}}{(a_{11}^2 b_{11}^2 - c_{11}^2)} \right] - \sum_{h=1}^{L} P_h^2 [A_h + R_h^2 B_h - 2R_h C_h] < 0
   \]

The proposed class of estimators is more efficient than other existing estimators when above conditions 1 to 4 are satisfied.

Numerical results

In this section three populations are generated for simulation study and four are based on real data sets. The results are given in Tables 1–3 (simulation) and 4–7 (real data).

Simulation study

We have generated three populations from normal distribution by using R language program. First population is generated for equal strata and the second one is generated for unequal strata and third one is generated for equal strata of small sample size (see Appendix).

The above tables shows that a general class of proposed estimators outperform all the other existing estimators. For \( \alpha = 0 \), \( \bar{y}_{S(P)} \) shows the better performance.

Application to real data

In this section we have considered four data sets (see Appendix) for numerical comparisons and results are given in Tables 4–7.

---

Table 1. Mean squared error (MSE) of different estimators for Pop I(N = 4000).

| Estimators | 10% non-response | 20% non-response |
|------------|-----------------|-----------------|
| \( \bar{y}_{S(HH)} \) | 0.095391 | 0.116525 | 0.158794 | 0.105952 | 0.148191 | 0.232694 |
| \( \bar{y}_{S(A)} \) | 0.021870 | 0.026292 | 0.035128 | 0.024495 | 0.034155 | 0.053486 |
| \( \bar{y}_{S(D)} \) | 0.020585 | 0.024792 | 0.033190 | 0.023020 | 0.032094 | 0.050240 |
| \( \bar{y}_{S(AH)} \) | 0.056410 | 0.068214 | 0.091802 | 0.063076 | 0.088209 | 0.138450 |
| \( \alpha = 0 \), \( \bar{y}_{S(P)} \) | 0.020450 | 0.024599 | 0.032852 | 0.022853 | 0.031779 | 0.049480 |
| \( \alpha = 1 \), \( \bar{y}_{S(P)} \) | 0.020458 | 0.024609 | 0.032873 | 0.022862 | 0.031797 | 0.049526 |
| \( \alpha = -1 \), \( \bar{y}_{S(P)} \) | 0.020459 | 0.024610 | 0.032874 | 0.022863 | 0.031798 | 0.049529 |

https://doi.org/10.1371/journal.pone.0191572.t001
| Estimators | 10% non-response | 20% non-response |
|-----------|-----------------|-----------------|
|           | g₀               | g₀              |
|           | 2    | 4    | 8    | 2    | 4    | 8    |
| YIH       | 0.094074        | 0.114968        | 0.156755 | 0.102687 | 0.140807 | 0.217047 |
| YIR       | 0.012512        | 0.015799        | 0.022374 | 0.014212 | 0.020899 | 0.034275 |
| YR(R)     | 0.012504        | 0.015776        | 0.022302 | 0.014193 | 0.020816 | 0.034010 |
| YR(S)     | 0.027283        | 0.034279        | 0.048270 | 0.030751 | 0.044685 | 0.072551 |
| a = 0, YR(P) | 0.012463    | 0.015712        | 0.022175 | 0.014141 | 0.020707 | 0.033732 |
| a = 1, YR(P) | 0.012466   | 0.015716        | 0.022182 | 0.014144 | 0.020713 | 0.033738 |
| a = -1, YR(P) | 0.012467  | 0.015717        | 0.022182 | 0.014145 | 0.020714 | 0.033739 |

https://doi.org/10.1371/journal.pone.0191572.t002

| Estimators | 10% non-response | 20% non-response |
|-----------|-----------------|-----------------|
|           | g₀               | g₀              |
|           | 2    | 4    | 8    | 2    | 4    | 8    |
| YIH       | 0.905181        | 1.034025        | 1.291698 | 1.004663 | 1.332462 | 1.988063 |
| YIR       | 0.280667        | 0.331342        | 0.432703 | 0.302164 | 0.395841 | 0.583195 |
| YR(R)     | 0.199208        | 0.232876        | 0.299860 | 0.212367 | 0.272204 | 0.391440 |
| YR(S)     | 0.559174        | 0.655443        | 0.847981 | 0.614328 | 0.820925 | 1.234103 |
| a = 0, YR(P) | 0.193840    | 0.225546        | 0.287533 | 0.206202 | 0.261988 | 0.370120 |
| a = 1, YR(P) | 0.193986   | 0.225734        | 0.287820 | 0.206378 | 0.262264 | 0.370868 |
| a = -1, YR(P) | 0.194009  | 0.225766        | 0.287885 | 0.206396 | 0.262314 | 0.370836 |

https://doi.org/10.1371/journal.pone.0191572.t003

| Estimators | 10% non-response | 20% non-response |
|-----------|-----------------|-----------------|
|           | g₀               | g₀              |
|           | 2    | 4    | 8    | 2    | 4    | 8    |
| YIH       | 2719483          | 3340426         | 4582312 | 3025218 | 4257631 | 6722457 |
| YIR       | 435763.4         | 536225.8        | 737150.5 | 470462.5 | 640323.1 | 980044.2 |
| YR(R)     | 278381.2         | 341063.4        | 466350.4 | 298383.2 | 400621.3 | 601701.3 |
| YR(S)     | 293568.4         | 360869.7        | 495472.4 | 316102.2 | 428471.3 | 653209.4 |
| a = 0, YR(P) | 232820.0     | 276450.6        | 356353.0 | 247489.8 | 316893.5 | 438483.9 |
| a = 1, YR(P) | 233140.6    | 276830          | 356763.9 | 247852.6 | 317379.8 | 438961.3 |
| a = -1, YR(P) | 233433.5   | 277328.6        | 357873.8 | 248211.2 | 318171.9 | 441257.5 |

https://doi.org/10.1371/journal.pone.0191572.t004
Table 5. Mean squared error (MSE) of different estimators for Pop V(N = 120).

| Estimators | 10% non-response | 20% non-response |
|------------|------------------|------------------|
|            | H0               | H0               |
|            | 2                | 4                | 8                | 2                | 4                | 8                |
| $\gamma_{Hh}$ | 219.5693 | 280.1731 | 401.3808 | 254.9066 | 386.1849 | 648.7416 |
| $\gamma_{R(o)}$ | 311.3524 | 395.5183 | 563.8503 | 304.1427 | 373.8894 | 513.3828 |
| $\gamma_{S(o)}$ | 191.9500 | 243.8635 | 347.6900 | 218.4376 | 316.3906 | 489.3592 |
| $\gamma_{AH}$ | 1041.79 | 1330.72 | 1908.6000 | 932.9254 | 1004.13 | 1146.5440 |
| $\gamma_{P}$ | a = 0 | $\gamma_{P}$ | a = 1 | $\gamma_{P}$ | a = 1 | $\gamma_{P}$ | a = 1 |

https://doi.org/10.1371/journal.pone.0191572.t005

Table 6. Mean squared error (MSE) of different estimators for Pop VI(N = 120).

| Estimators | 10% non-response | 20% non-response |
|------------|------------------|------------------|
|            | H0               | H0               |
|            | 2                | 4                | 8                | 2                | 4                | 8                |
| $\gamma_{Hh}$ | 249.1903 | 369.0362 | 608.728 | 254.9060 | 386.1843 | 648.741 |
| $\gamma_{R(o)}$ | 0.400780 | 0.594000 | 0.980242 | 1.921332 | 5.155440 | 11.6236 |
| $\gamma_{S(o)}$ | 0.354179 | 0.5248469 | 0.8659781 | 1.79562 | 4.655160 | 10.1316 |
| $\gamma_{AH}$ | 352.2648 | 522.7970 | 863.8607 | 363.5490 | 556.6457 | 942.3837 |
| $\gamma_{P}$ | a = 0 | $\gamma_{P}$ | a = 1 | $\gamma_{P}$ | a = 1 | $\gamma_{P}$ | a = 1 |

https://doi.org/10.1371/journal.pone.0191572.t006

Table 7. Mean squared error (MSE) of different estimators for Pop VII(N = 119).

| Estimators | 10% non-response | 20% non-response |
|------------|------------------|------------------|
|            | H0               | H0               |
|            | 2                | 4                | 8                | 2                | 4                | 8                |
| $\gamma_{Hh}$ | 630.3935 | 805.3869 | 1155.3740 | 713.2246 | 1053.8800 | 1735.1910 |
| $\gamma_{R(o)}$ | 55.4937 | 70.4183 | 100.2675 | 61.3649 | 88.0319 | 141.3660 |
| $\gamma_{S(o)}$ | 54.9625 | 69.7884 | 99.4359 | 60.7277 | 87.0779 | 139.7685 |
| $\gamma_{AH}$ | 508.7321 | 652.4944 | 940.0190 | 574.8255 | 850.7745 | 1402.6730 |
| $\gamma_{P}$ | a = 0 | $\gamma_{P}$ | a = 1 | $\gamma_{P}$ | a = 1 | $\gamma_{P}$ | a = 1 |

https://doi.org/10.1371/journal.pone.0191572.t007
In these tables, we observed that a general class of proposed estimators is more efficient as compared to all other considered estimators. For $\alpha = 0$, $\bar{y}_{SP}$ shows the better performance.

**Conclusion**

In the present paper, we have suggested an improved class of estimators of the finite population mean in the presence of measurement error and non-response under stratified random sampling. Through simulation study and real life data sets it is observed that the proposed class of estimators perform better than the existing estimators. The mean square error values are generally smaller under 10% of non-response as compared to 20% of non-response, which are expected results. Generally as the non-response rate increases, mean square error also increases. Based on numerical findings, it turns out that the proposed class of estimators is more efficient for the situations when $\alpha = 0$, $\alpha = 1$ and $\alpha = -1$ as compared to the other existing estimators. Among different classes, the performance of proposed class of estimators is better for $\alpha = 0$ in Tables 1–7.

**Appendix**

**Population I.**

$X_1 = rnorm(1000, 5, 10)$, $Y_1 = X_1 + rnorm(1000, 0, 1)$, $y_1 = Y_1 + rnorm(1000, 1, 3)$, $x_1 = X_1 + rnorm(1000, 1, 3)$

$X_2 = rnorm(1000, 4, 8)$, $Y_2 = X_2 + rnorm(1000, 0, 1)$, $y_2 = Y_2 + rnorm(1000, 1, 3)$, $x_2 = X_2 + rnorm(1000, 1, 3)$

$X_3 = rnorm(1000, 4, 9)$, $Y_3 = X_3 + rnorm(1000, 0, 1)$, $y_3 = Y_3 + rnorm(1000, 1, 3)$, $x_3 = X_3 + rnorm(1000, 1, 3)$

$X_4 = rnorm(1000, 3, 7)$, $Y_4 = X_4 + rnorm(1000, 0, 1)$, $y_4 = Y_4 + rnorm(1000, 1, 3)$, $x_4 = X_4 + rnorm(1000, 1, 3)$

Number of Stratas = 4

$N_1 = 1000$, $N_2 = 1000$, $N_3 = 1000$, $N_4 = 1000$, $n_1 = 200$, $n_2 = 200$, $n_3 = 200$, $n_4 = 200$

$\bar{Y}_1 = 5.670898$, $\bar{Y}_2 = 3.646459$, $\bar{Y}_3 = 4.039068$, $\bar{Y}_4 = 2.902825$, $\bar{X}_1 = 5.666893$

$\bar{X}_1 = 3.643237$, $\bar{X}_2 = 3.968049$, $\bar{X}_3 = 2.918596$, $S_{XY} = 106.310$, $S_{Xy} = 67.69053$

$S_{ux} = 81.46952$, $S_{yx} = 47.2278$, $S_{ux} = 104.2774$, $S_{xy} = 66.19725$, $S_{yx} = 81.06883$

$S_{ix} = 45.99937$, $\rho_{1YX} = 0.9950779$, $\rho_{2YX} = 0.9926346$, $\rho_{3YX} = 0.9939164$, $\rho_{4YX} = 0.9896319$

**Population II.**

$X_1 = rnorm(1000, 5, 10)$, $Y_1 = X_1 + rnorm(1000, 0, 1)$, $y_1 = Y_1 + rnorm(1000, 1, 3)$, $x_1 = X_1 + rnorm(1000, 0, 1)$

$X_2 = rnorm(1200, 4, 8)$, $Y_2 = X_2 + rnorm(1200, 0, 1)$, $y_2 = Y_2 + rnorm(1200, 1, 3)$, $x_2 = X_2 + rnorm(1200, 0, 1)$

$X_3 = rnorm(1300, 4, 9)$, $Y_3 = X_3 + rnorm(1300, 0, 1)$, $y_3 = Y_3 + rnorm(1300, 1, 3)$, $x_3 = X_3 + rnorm(1300, 0, 1)$

$X_4 = rnorm(1500, 3, 7)$, $Y_4 = X_4 + rnorm(1500, 0, 1)$, $y_4 = Y_4 + rnorm(1500, 1, 3)$, $x_4 = X_4 + rnorm(1500, 1, 3)$

Number of Stratas = 4

$N_1 = 1000$, $N_2 = 1200$, $N_3 = 1300$, $N_4 = 1500$, $n_1 = 200$, $n_2 = 210$, $n_3 = 220$, $n_4 = 215$

$\bar{Y}_1 = 5.670898$, $\bar{Y}_2 = 3.79172$, $\bar{Y}_3 = 4.582486$, $\bar{Y}_4 = 3.20716$, $\bar{X}_1 = 5.666893$

$\bar{X}_1 = 3.807569$, $\bar{X}_2 = 4.627208$, $\bar{X}_3 = 3.241139$, $S_{xy} = 106.310$, $S_{xy} = 66.3915$

$S_{ux} = 84.23239$, $S_{ux} = 53.52266$, $S_{ux} = 104.2774$, $S_{ux} = 65.46337$, $S_{ux} = 82.98812$

$S_{ux} = 52.84269$, $\rho_{1YX} = 0.9950779$, $\rho_{2YX} = 0.9924618$, $\rho_{3YX} = 0.9939475$, $\rho_{4YX} = 0.9903624$. 
Population III.

\[ X_1 = \text{rnorm}(200, 5, 10), \ y_1 = X_1 + \text{rnorm}(200, 0, 1), \ x_1 = X_1 + \text{rnorm}(200, 0, 1) \]
\[ X_2 = \text{rnorm}(200, 4, 8), \ y_2 = X_2 + \text{rnorm}(200, 0, 1), \ x_2 = X_2 + \text{rnorm}(1200, 0, 1) \]
\[ X_3 = \text{rnorm}(200, 4, 9), \ y_3 = X_3 + \text{rnorm}(200, 0, 1), \ x_3 = X_3 + \text{rnorm}(1300, 0, 1) \]
\[ X_4 = \text{rnorm}(200, 3, 7), \ y_4 = X_4 + \text{rnorm}(200, 0, 1), \ x_4 = X_4 + \text{rnorm}(200, 1, 3) \]

Number of Stratas = 4

\[ N_1 = 200, \ N_2 = 200, \ N_3 = 200, \ N_4 = 200, \ n_1 = 22, \ n_2 = 28, \ n_3 = 22, \ n_4 = 25, \ Y_1 = 7.470477, \]
\[ Y_2 = 4.815927, \ Y_3 = 5.206734, \ Y_4 = 4.864139, \bar{X}_1 = 5.895905, \bar{X}_2 = 3.720469, \]
\[ \bar{X}_3 = 3.9945, \bar{X}_4 = 3.698503, S_{1y} = 113.6437, S_{2y} = 73.19005, S_{3y} = 92.51202, \]
\[ S_{4y} = 62.38946, S_{1x}^2 = 105.978, S_{2x}^2 = 68.09821, S_{3x}^2 = 89.87238, \rho_{1yx} = 0.9447295, \rho_{2yx} = 0.9214058, \rho_{3yx} = 0.9529337, \rho_{4yx} = 0.912232. \]

Population IV. (Source: [18])

\( Y \): Number of teachers, \( X \): Number of students.

Number of Stratas = 6

\[ N_1 = 106, \ N_2 = 106, \ N_3 = 94, \ N_4 = 171, \ N_5 = 204, \ N_6 = 173, \ n_1 = 15, \ n_2 = 15, \ n_3 = 12, \ n_4 = 20, \]
\[ n_5 = 23, \ n_6 = 15, \ Y_1 = 1536.774, \ Y_2 = 2212.594, \ Y_3 = 9384.309, \ Y_4 = 5588.012, \]
\[ \bar{Y}_1 = 966.9559, \bar{Y}_2 = 404.3988, \bar{X}_1 = 2437.59, \bar{X}_2 = 27421.70, \bar{X}_3 = 72409.95, \]
\[ \bar{X}_4 = 74364.68, \bar{X}_5 = 26441.72, \bar{X}_6 = 9842.15, S_{1y} = 41281746, S_{2y} = 133437791, \]
\[ S_{3y} = 894457433, S_{4y}^2 = 82045636, S_{5y}^2 = 5710999, S_{6y}^2 = 894440.3, S_{1x}^2 = 2419565835, \]
\[ S_{2x}^2 = 330172268, S_{3x}^2 = 25842911895, S_{4x}^2 = 8156914688, S_{5x}^2 = 2061412416, \]
\[ S_{6x}^2 = 353245374, \rho_{1yx} = 0.8156414, \rho_{2yx} = 0.8156414, \rho_{3yx} = 0.9011201, \rho_{4yx} = 0.9858761, \rho_{5yx} = 0.7130988, \rho_{6yx} = 0.893595. \]

Population V. (Source: [19])

\( Y \): 1983 population(in millions), \( X \): 1982 gross national product

Number of Stratas = 5

\[ N_1 = 38, \ N_2 = 14, \ N_3 = 11, \ N_4 = 33, \ N_5 = 24, \ n_1 = 17, \ n_2 = 6, \ n_3 = 4, \ n_4 = 12, \ n_5 = 11, \]
\[ \bar{Y}_1 = 13.03684, \bar{Y}_2 = 27.35, \bar{Y}_3 = 23.13636, \bar{Y}_4 = 79.65455, \bar{Y}_5 = 20.28333, \]
\[ \bar{X}_1 = 1029.158, \bar{X}_2 = 25671.57, \bar{X}_3 = 5028.818, \bar{X}_4 = 7533.939, \bar{X}_5 = 16315.25. \]
\[ S_{1y}^2 = 270.9083, S_{2y}^2 = 3906.929, S_{3y}^2 = 1339.405, S_{4y}^2 = 45082.17, S_{5y}^2 = 368.9423, \]
\[ S_{1x}^2 = 3667896, S_{2x}^2 = 6568461403, S_{3x}^2 = 63348743, S_{4x}^2 = 440717912, S_{5x}^2 = 408412121, \]
\[ \rho_{1yx} = 0.7439544, \rho_{2yx} = 0.969956, \rho_{3yx} = 0.9768227, \rho_{4yx} = 0.2948897, \rho_{5yx} = 0.9011072. \]

Population VI. (Source: [19])

\( Y \): 1983 population(in millions), \( X \): 1980 population(in millions)

Number of Stratas = 5

\[ N_1 = 38, \ N_2 = 14, \ N_3 = 11, \ N_4 = 33, \ N_5 = 24, \ n_1 = 17, \ n_2 = 6, \ n_3 = 4, \ n_4 = 12, \ n_5 = 11, \]
\[ \bar{Y}_1 = 13.03684, \bar{Y}_2 = 27.35, \bar{Y}_3 = 23.13636, \bar{Y}_4 = 79.65455, \bar{Y}_5 = 20.28333, \]
\[ \bar{X}_1 = 11.88421, \bar{X}_2 = 26.18571, \bar{X}_3 = 21.88182, \bar{X}_4 = 75.24242, \bar{X}_5 = 20.09583, \]
\[ S_{1y}^2 = 270.9083, S_{2y}^2 = 3906.929, S_{3y}^2 = 1339.405, S_{4y}^2 = 45082.17, S_{5y}^2 = 368.9423, \]
\[ S_{1x}^2 = 222.4889, S_{2x}^2 = 3683.071, S_{3x}^2 = 1174.032, S_{4x}^2 = 41280.19, S_{5x}^2 = 364.7839, \rho_{1yx} = 0.9996193, \rho_{2yx} = 0.9998693, \rho_{3yx} = 0.9998858, \rho_{4yx} = 0.9993071, \rho_{5yx} = 0.9998059. \]
Population VII. ([20])

\[ Y: \] Production of wheat (in tons), \( X: \) Area of wheat (in hectares)

Number of Stratas = 4

\[ N_1 = 47, N_2 = 30, N_3 = 29, N_4 = 13, n_1 = 15, n_2 = 10, n_3 = 10, n_4 = 5, \bar{Y}_1 = 443.5447, \]
\[ \bar{Y}_2 = 68.68276, \bar{Y}_3 = 17.06667, \bar{Y}_4 = 52.52308, \bar{X}_1 = 160.2362, \bar{X}_2 = 29.70345, \]
\[ \bar{X}_3 = 11.54667, \bar{X}_4 = 23.62308, S^2_{1Y} = 74026.75, S^2_{2Y} = 2871.781, S^2_{3Y} = 244.1292, \]
\[ S^2_{4Y} = 4451.124, S^2_{1X} = 8377.401, S^2_{2X} = 316.4532, S^2_{3X} = 91.45775, S^2_{4X} = 682.9703, \rho_{1YX} = 0.9583838, \rho_{2YX} = 0.779071, \rho_{3YX} = 0.8719665, \rho_{4YX} = 0.9922591. \]

Supporting information

S1 File. Data used in the manuscript “S1 File.zip”.

(ZIP)

Acknowledgments

The authors are grateful to the two anonymous referees for their valuable comments and feedback.

Author Contributions

Conceptualization: Erum Zahid.
Data curation: Erum Zahid.
Formal analysis: Erum Zahid.
Funding acquisition: Erum Zahid.
Investigation: Erum Zahid, Javid Shabbir.
Methodology: Erum Zahid.
Project administration: Javid Shabbir.
Resources: Erum Zahid, Javid Shabbir.
Software: Erum Zahid.
Supervision: Javid Shabbir.
Validation: Erum Zahid, Javid Shabbir.
Visualization: Erum Zahid, Javid Shabbir.
Writing – original draft: Erum Zahid.
Writing – review & editing: Erum Zahid.

References

1. Hansen MH, Hurwitz WN. The problem of non-response in sample surveys. Journal of the American Statistical Association. 1946; 41(236):517–529. https://doi.org/10.1080/01621459.1946.10501894

PMID: 20279350

2. Cochran WG. Sampling techniques. New York: John Wills & Sons. 1977;

3. Rao P. Ratio estimation with subsampling the nonrespondents. Survey Methodology. 1986; 12(2):217–230.

4. Khare B, Srivastava S. Generalized two phase sampling estimators for the population mean in the presence of nonresponse. Alligarh Journal of Statistics. 2010; 30:39–54.
5. Singh HP, Kumar S, et al. Combination of regression and ratio estimate in presence of nonresponse. Brazilian journal of Probability and Statistics. 2011; 25(2):205-217. https://doi.org/10.1214/10-BJPS117

6. Shabbir J, Khan NS. Some Modified Exponential-Ratio Type Estimators in the presence of Non-response under Two-Phase Sampling Scheme. Electronic Journal of Applied Statistical Analysis. 2013; 6(1):1-17.

7. Cochran WG. Errors of measurement in statistics. Technometrics. 1968; 10(4):637-666. https://doi.org/10.2307/1267450

8. Fuller WA. Estimation in the presence of measurement error. International Statistical Review/Revue Internationale de Statistique. 1995; p. 121–141.

9. Shalabh S. Ratio method of estimation in the presence of measurement errors. Jour Ind Soc Agri Statist. 1997; 52:150–155.

10. Biemer PP, Groves RM, Lyberg LE, Mathiowetz NA, Sudman S. Measurement Errors in Surveys. John Wiley & Sons; 2011.

11. 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):1–8.

12. Kumar S, Bhougal S, Nataraja N, Viswanathaiah M. Estimation of Population Mean in the Presence of Non-Response and Measurement Error. Revista Colombiana de Estadistica. 2015; 38(1):145–161. https://doi.org/10.15446/rce.v38n1.49807

13. Singh RS, Sharma P. Method of Estimation in the Presence of Non-response and Measurement Errors Simultaneously. Journal of Modern Applied Statistical Methods. 2015; 14(1):12. https://doi.org/10.22237/jmasm/1430453460

14. Azeem M, Hanif M. Joint influence of measurement error and non response on estimation of population mean. Communications in Statistics-Theory and Methods. 2017; 46(4):1679–1693. https://doi.org/10.1080/03610926.2015.1026992

15. Kumar S. Improved estimation of population mean in presence of nonresponse and measurement error. Journal of Statistical Theory and Practice. 2016; 10(4):707–720. https://doi.org/10.1080/15598608.2016.121648

16. Shabbir J, Gupta S. On estimating finite population mean in simple and stratified random sampling. Communications in Statistics-Theory and Methods. 2010; 40(2):199–212. https://doi.org/10.1080/0361092093411259

17. Haq A, Shabbir J. Improved family of ratio estimators in simple and stratified random sampling. Communications in Statistics-Theory and Methods. 2013; 42(5):782–799. https://doi.org/10.1080/03610926.2011.579377

18. Koyuncu N, Kadilar C. Ratio and product estimators in stratified random sampling. Journal of Statistical Planning and Inference. 2009; 139(8):2552–2558. https://doi.org/10.1016/j.jspi.2008.11.009

19. Sarmid C, Swensson B, Wretman J. Model Assisted Survey Sampling. New York: Springer. 1992.

20. FBS. Crops area production by districts, Islamabad; 2011.