Almost Unbiased Estimation of Coefficient of Dispersion from Incomplete Data

Muhammad Umair Sohail¹*, Nursel Koyuncu², and Muhammad Areeb Iqbal Sethi³

¹Department of Statistics, University of Narowal, Narowal, Pakistan
²Department of Statistics, Hacettepe University, Beyete, Ankara, Turkey
³Department of Management Sciences, University of Narowal, Narowal, Pakistan

*Corresponding Author: umairsohailch@gmail.com

Abstract

This article develops an almost unbiased estimation of coefficient of dispersion by the productive use of coefficient of dispersion of the auxiliary variable in two phase sampling. Expressions for variances of the proposed estimators are obtained up to first order of approximation. The relative comparison of proposed unbiased ratio estimator are compared with naive estimator by using simulated data sets. Thus, we conclude that the suggested imputation methodology is more efficient than traditional estimator.

Keywords: Auxiliary variable, Coefficient of dispersion, Missing data, Relative efficiency.

Introduction

In last few decades, the problem of estimating the population parameters such as mean, median, variance, and correlation coefficient are significantly considered by using the suitable auxiliary information either at the design stage, estimation stage, or both stages. The precise estimation of population parameters has got lot of attention with the productive use of auxiliary information. Whole estimation procedure is quite laborious and inefficient, when the observed/sampled data have missing values. These missing values can spoil the entire estimation procedure.

Hansen and Hurwitz (1946) was the first, who considered the problem of non-response in telephone survey by dividing the population in two groups, says: respondent group having \( N_1 \) units and non-respondent group having \( N_2 \) units. Later on, Rubin (1976) defined the different mechanism for non-response are: (i) missing at random (MAR), (ii) observed at random (OAR), and (iii) missing complete at random (MCAR). For the sake of precise and valid inference about the parameters of interest, imputation is reliable procedure to impute/suggest the missing values and remove the non-response bias. Rubin (1978) provided the idea of multiple imputation with significant use of available information. Furthermore, Little and Rubin (1987), Sarndal (1992), Heitjan and Basu (1996), Singh and Horn (2000), Singh and Deo (2003), Rubin (2004), Longford
(2006), and many others provide the detailed discussion on type and nature of missing structure. Singh et al. (1996) use the Quenouille’s method for the almost unbiased estimation in double sampling design.

Motivated by Singh and Singh (1993) and Ambati et al. (2018) estimation procedures, we consider the point estimation of coefficient of dispersion in two phase sampling design for the case of missing values. Let $Y_1, Y_2, Y_3, \cdots, Y_N$ be the values of the study variable in a finite population with mean $\bar{Y} = \frac{1}{N} \sum_{j=1}^{N} Y_j$ and variance $S_Y^2 = \frac{1}{(N-1)} \sum_{j=1}^{N} (Y_j - \bar{Y})^2$. Let, we have a random sample (s’) of size $n’$ is drawn at first phase and in second phase, sample (s) of $n$ units are selected. In the present study, we considered that the non-response in occurred only in the study variable. Out of $n$ units, only $n_1$ units provide the response regarding the characteristics of interest, belongs to group $G$ and remaining $n_2 = (n - n_1)$ units are non-respondents, belongs to group $G^c$. So, the respondent group have the mean $\bar{y}_{n_1} = \frac{1}{n_1} \sum_{j=1}^{n_1} y_j$.

The rest of the article is designed as follows: In Section 2, we revisit to some traditional imputation procedures for the imputation of missing values in the estimation of population coefficient of dispersion. In Section 3, we derive large sample approximations of $Q_1$ and $Q_3$. In Section 4, we proposed asymptotically unbiased procedure is for the imputation of missing values. The numerical results of simulation study are reported in Section 5. In section 6, we conclude our study.

**Review of Literature**

Ambati et al. (2018) considered the estimation of coefficient of dispersion in simple random sampling, as, let $\hat{Q}_{1(y)}$ and $\hat{Q}_{3(y)}$ are the estimator of $Q_{1(y)}$ and $Q_{3(y)}$, respectively, of the study variable and also let $\hat{Q}_{1(x)}$ and $\hat{Q}_{3(x)}$ are the estimator of $Q_{1(x)}$ and $Q_{3(x)}$ respectively of the auxiliary variable. The population coefficient of quartile deviation for $Y$ and $X$ are define as:

\[
CQD_y = \frac{Q_{3(y)} - Q_{1(y)}}{Q_{3(y)} + Q_{1(y)}} \quad \text{and} \quad CQD_x = \frac{Q_{3(x)} - Q_{1(x)}}{Q_{3(x)} + Q_{1(x)}}
\]

(i) Rewrite the coefficient of quartile deviation for the situation of missing values in two phase sampling, as:

\[
\Delta_{1j} = \begin{cases} 
  y_j & \text{if } j \in G \\
  \hat{CQD}_y(n_1) & \text{if } j \in G^c
\end{cases}
\]

where $\hat{CQD}_y(n_1) = \frac{\hat{Q}_{3(y)(n_1)} - \hat{Q}_{1(y)(n_1)}}{\hat{Q}_{3(y)(n_1)} + \hat{Q}_{1(y)(n_1)}}$. The point estimator for population coefficient of dispersion is given by

\[
\hat{CQD}_{\text{naive}} = \hat{CQD}_y(n_1) = \frac{\hat{Q}_{3(y)(n_1)} - \hat{Q}_{1(y)(n_1)}}{\hat{Q}_{3(y)(n_1)} + \hat{Q}_{1(y)(n_1)}}
\]

Bias and mean square error
Almost Unbiased Estimation of Coefficient of Dispersion from Incomplete Data

\[ \text{Bias}(\hat{CQD}_{naive}) \approx \Pi_{(n_1,n)} CQD^2 \left\{ \frac{3}{16} \left\{ \frac{1}{Q_3(y)(f(Q_3(y)))^2} + \frac{1}{Q_1(y)(f(Q_1(y)))^2} \right\} + \left( \frac{2}{3} \right) \left( \frac{1}{Q_1(y), Q_3(y) \{f(Q_1(y)) f(Q_3(y))\}} \right) \right\} - \left( \frac{3}{16} \right) \left( \frac{1}{Q_3(y)(f(Q_3(y)))^2} + \frac{1}{Q_1(y)(f(Q_1(y)))^2} \right) \] \]

and

\[ \text{MSE}(\hat{CQD}_{naive}) \approx \Pi_{(n_1,n)} CQD^2 \left\{ \frac{1}{(Q_3(y)-Q_1(y))} + \frac{1}{(Q_3(y)+Q_1(y))} \right\}^2 \left\{ \frac{3}{16} \left\{ \frac{1}{Q_3(y)(f(Q_3(y)))^2} + \frac{1}{Q_1(y)(f(Q_1(y)))^2} \right\} \right\} - \left( \frac{2}{3} \right) \left( \frac{1}{Q_1(y), Q_3(y) \{f(Q_1(y)) f(Q_3(y))\}} \right) \] \]

(ii) Rewriting the proposed estimator by Cochran (1940) for the case of coefficient of quartile deviation is given by as:

\[ \Delta_{2j} = \left\{ \frac{1}{1-f} \left( \hat{CQD}_{y(n_1)} - f \hat{CQD}_{y(n_1)} \right) \right\} \] if \( j \in G \)

where \( f = \frac{n_1}{n} \) \( \hat{CQD}_{x(n)} = \frac{\hat{Q}_3[x(n)] - \hat{Q}_1[x(n)]}{\hat{Q}_3[x(n)] + \hat{Q}_1[x(n)]} \) and \( \hat{CQD}_{x(n_1)} = \frac{\hat{Q}_3[x(n_1)] - \hat{Q}_1[x(n_1)]}{\hat{Q}_3[x(n_1)] + \hat{Q}_1[x(n_1)]} \). The point estimator is defined as follow:

\[ \hat{CQD}_R = \hat{CQD}_{y(n_1)} \frac{\hat{CQD}_{x(n)}}{\hat{CQD}_{x(n_1)}} \]

Mean square error is given by

\[ \text{MSE}(\hat{CQD}_{naive}) \approx \Pi_{(n_1,n)} CQD^2 \left\{ \frac{1}{(Q_3(y)-Q_1(y))} + \frac{1}{(Q_3(y)+Q_1(y))} \right\}^2 \left\{ \frac{3}{16} \left\{ \frac{1}{Q_3(y)(f(Q_3(y)))^2} + \frac{1}{Q_1(y)(f(Q_1(y)))^2} \right\} \right\} - \left( \frac{2}{3} \right) \left( \frac{1}{Q_1(y), Q_3(y) \{f(Q_1(y)) f(Q_3(y))\}} \right) \] \]

\[ + \Pi_{(n,n)} CQD^2 \left\{ \frac{1}{(Q_3(x) - Q_1(x))} + \frac{1}{(Q_3(x) + Q_1(x))} \right\}^2 \left\{ \frac{3}{16} \left\{ \frac{1}{Q_3(x)(f(Q_3(x)))^2} + \frac{1}{Q_1(x)(f(Q_1(x)))^2} \right\} \right\} - \left( \frac{2}{3} \right) \left( \frac{1}{Q_1(x), Q_3(x) \{f(Q_1(x)) f(Q_3(x))\}} \right) \]
\[ -2\Pi_{(n_1,n')} CQD^2 \left( \frac{1}{(\mathcal{Q}_3(y)-\mathcal{Q}_1(y))} - \frac{1}{(\mathcal{Q}_3(y)+\mathcal{Q}_1(y))} \right) \left( \frac{1}{\mathcal{Q}_3(x)-\mathcal{Q}_2(x)} \right) \]

\[ - \frac{1}{(\mathcal{Q}_3(x)+\mathcal{Q}_1(x))} \left( \frac{\mathcal{P}_Q(y)\mathcal{Q}_3(x)}{\mathcal{f}_Q(y)\mathcal{f}_Q(x)} - \frac{\mathcal{P}_Q(\mathcal{Q}_1(y)\mathcal{Q}_3(x))}{\mathcal{f}_Q(y)\mathcal{f}_Q(x)} \right) \]

\[ - \frac{\mathcal{P}_Q(y)\mathcal{Q}_1(x)}{\mathcal{f}_Q(y)\mathcal{f}_Q(x)} \left( \frac{\mathcal{P}_Q(\mathcal{Q}_1(y)\mathcal{Q}_3(x))}{\mathcal{f}_Q(y)\mathcal{f}_Q(x)} + \frac{1}{16} \right) \]

(6)

**Notations and Expectations**

On the lines of Ambati et al. (2018), we consider the following probability table as:

|        | \( Y \leq \mathcal{Q}_1(y) \) | \( Y \leq \mathcal{Q}_3(y) \) | \( X \leq \mathcal{Q}_1(x) \) | \( X \leq \mathcal{Q}_3(x) \) |
|--------|----------------|----------------|----------------|----------------|
| \( Y \leq \mathcal{Q}_1(y) \) | 0.25           |               |               |               |
| \( Y \leq \mathcal{Q}_3(y) \) |               | 0.75          |               |               |
| \( X \leq \mathcal{Q}_1(x) \) | \( P \) \( q_1(y)q_1(x) \) | \( P \) \( q_3(y)q_1(x) \) | 0.25          |               |
| \( X \leq \mathcal{Q}_3(x) \) | \( P \) \( q_1(y)q_3(x) \) | \( P \) \( q_3(y)q_3(x) \) | 0.25          | 0.75          |

For the mathematical expressions of bias and mean square error, let

\( \zeta_1 = \frac{\mathcal{Q}_1(y)}{\mathcal{Q}_1(y)} - 1 \), \( \zeta_2 = \frac{\mathcal{Q}_3(y)}{\mathcal{Q}_3(y)} - 1 \), \( \zeta_3 = \frac{\mathcal{Q}_1(x)}{\mathcal{Q}_1(x)} - 1 \), \( \zeta_4 = \frac{\mathcal{Q}_3(x)}{\mathcal{Q}_3(x)} - 1 \),

\( \zeta_5 = \frac{\mathcal{Q}_1(x)}{\mathcal{Q}_1(x)} - 1 \), \( \zeta_6 = \frac{\mathcal{Q}_3(x)}{\mathcal{Q}_3(x)} - 1 \), \( E(\zeta_i) = 0 \) for \( i = 1, 2, \cdots, \) and 6.

Up to first order approximation, we have

\[ E(\zeta_2) = \Pi_{(n_1,n')} \left( \frac{3}{16} \mathcal{Q}_1(y) \mathcal{f}(\mathcal{Q}_1(y)) \right)^2 \],

\[ E(\zeta_3) = \Pi_{(n,n')} \left( \frac{3}{16} \mathcal{Q}_3(y) \mathcal{f}(\mathcal{Q}_3(y)) \right)^2 \],

\[ E(\zeta_4) = \Pi_{(n,n')} \left( \frac{3}{16} \mathcal{Q}_1(x) \mathcal{f}(\mathcal{Q}_1(x)) \right)^2 \],

\[ E(\zeta_5) = \Pi_{(n,n')} \left( \frac{3}{16} \mathcal{Q}_3(x) \mathcal{f}(\mathcal{Q}_3(x)) \right)^2 \],

\[ E(\zeta_6) = \Pi_{(n,n')} \left( \frac{3}{16} \mathcal{Q}_3(x) \mathcal{f}(\mathcal{Q}_3(x)) \right)^2 \],

\[ E(\zeta_1) = \Pi_{(n_1,n')} \left( \frac{1}{16} \mathcal{Q}_1(y) \mathcal{Q}_3(y) \mathcal{f}(\mathcal{Q}_1(y)) \mathcal{f}(\mathcal{Q}_3(y)) \right)^2 \],

\[ E(\zeta_3) = \Pi_{(n,n')} \left( \frac{1}{16} \mathcal{Q}_1(y) \mathcal{Q}_1(x) \mathcal{f}(\mathcal{Q}_1(y)) \mathcal{f}(\mathcal{Q}_1(x)) \right)^2 \],

\[ E(\zeta_4) = \Pi_{(n,n')} \left( \frac{1}{16} \mathcal{Q}_3(y) \mathcal{Q}_3(x) \mathcal{f}(\mathcal{Q}_3(y)) \mathcal{f}(\mathcal{Q}_3(x)) \right)^2 \].
Almost Unbiased Estimation of Coefficient of Dispersion from Incomplete Data

\[
E(\zeta_2 \zeta_5) = \Pi_{(n,n')} \left( \frac{1}{16}(P_{Q_3(Y)Q_1(X)} - \frac{3}{16}(\frac{1}{Q_3(Y)Q_1(X)\{f(Q_3(Y))f(Q_1(X))\}}), \right.
\]

\[
E(\zeta_2 \zeta_6) = \Pi_{(n,n')} \left( \frac{1}{16}(P_{Q_3(Y)Q_3(X)} - \frac{9}{16}(\frac{1}{Q_3(Y)Q_3(X)\{f(Q_3(Y))f(Q_3(X))\}}), \right.
\]

\[
E(\zeta_3 \zeta_4) = \Pi_{(n,n')} \left( \frac{1}{16}(\frac{1}{Q_1(X)Q_3(X)\{f(Q_1(X))f(Q_3(X))\}}), \right.
\]

\[
E(\zeta_5 \zeta_6) = \Pi_{(n,n')} \left( \frac{1}{16}(\frac{1}{Q_1(X)Q_3(X)\{f(Q_1(X))f(Q_3(X))\}}), \right.
\]

\[
E(\zeta_3 \zeta_5) = \Pi_{(n,n')} \left( \frac{1}{16}(\frac{1}{Q_1(X)Q_3(X)\{f(Q_1(X))f(Q_3(X))\}}), \right.
\]

\[
E(\zeta_4 \zeta_6) = \Pi_{(n,n')} \left( \frac{1}{16}(\frac{1}{Q_3(Y)Q_1(X)\{f(Q_3(Y))f(Q_1(X))\}}), \right.
\]

\[
E(\zeta_4 \zeta_5) = \Pi_{(n,n')} \left( \frac{1}{16}(\frac{1}{Q_3(Y)Q_1(X)\{f(Q_3(Y))f(Q_1(X))\}}), \right.
\]

\[
\text{where}
\]

\[
P_{Q_1(Y)Q_1(X)} = P(Y \leq Q_1(Y) \cap Y \leq Q_1(X)), \Pi_{(n,n')} = \left( \frac{1-f}{n} \right),
\]

\[
P_{Q_1(Y)Q_3(X)} = P(Y \leq Q_1(Y) \cap Y \leq Q_3(X)), \Pi_{(n,n')} = \left( \frac{1-f}{n} \right),
\]

\[
P_{Q_3(Y)Q_1(X)} = P(Y \leq Q_3(Y) \cap Y \leq Q_1(X)), \Pi_{(n,n')} = \left( \frac{1-f}{n} \right),
\]

\[
P_{Q_3(Y)Q_3(X)} = P(Y \leq Q_3(Y) \cap Y \leq Q_3(X)), f_1 = \frac{n}{N}, f_1 = \frac{n}{n}, f = \frac{n_1}{n}.
\]

In the next section, we consider an linear unbiased estimation of finite coefficient of dispersion in two phase sampling design.

Proposed Estimator for Coefficient of Dispersion

As we known from the available literature that the suitable auxiliary information not only helps at design stage and also improve the efficiency of the estimation procedure. Our main focus is to design an asymptotically unbiased estimator for the coefficient of depression by the significant use of the available auxiliary information. We derive a linear variety of ratio estimator which is unbiased in nature and have minimum variance among the class of ratio estimators. The proposed estimator is given by as:

\[
\hat{R}_i = \hat{C}QD_{\gamma(n)} \left( \frac{\hat{C}QD_{\gamma(n)}}{\hat{C}QD_{\gamma(n)}} \right)^i
\]

such that \( R_i \) \( \in \mathbb{C} \) for \( i = 1,2,3 \); where \( \mathbb{C} \) denote the class of all the possible linear ratio type estimator for estimating the population coefficient of dispersion (CQD\( _\gamma \)).

\[
\hat{R}_c = \sum_{i=1}^{3} c_i \hat{R}_i \text{ for } \sum_{i=1}^{3} c_i = 1 \text{ and } c_i \in \epsilon \in \mathbb{R}
\]

where \( c_i \) is the amount of antibiotic use to cure an estimator for biased and \( \mathbb{R} \) express the set of real number. Now, we extract the function \( \hat{R}_c \), so we have
\[
\hat{R}_c = \hat{C}QD_y(n_1)\{c_1\hat{R}_1 + c_2\hat{R}_2 + c_3\hat{R}_3\}
= \hat{C}QD_y(n_1)\left\{c_1 \cdot \left(\frac{\hat{C}QD_{x(n)}}{\hat{C}QD_{x(n)}}\right) + c_2 \cdot \left(\frac{\hat{C}QD_{x(n)}}{\hat{C}QD_{x(n)}}\right)^2 + c_3 \cdot \left(\frac{\hat{C}QD_{x(n)}}{\hat{C}QD_{x(n)}}\right)^3\right\}
\]

(9)

up to first order approximation, we can rewrite (4.3) in terms of error, then we solve (4.3) in parts, as:

\[
\hat{R}_1 = \frac{(\hat{C}QD_{3x(n)}) - \hat{C}QD_{1x(n)}}{\hat{C}QD_{3x(n)} + \hat{C}QD_{1x(n)}} \cdot \frac{(\hat{C}QD_{3x(n)}) - \hat{C}QD_{1x(n)}}{\hat{C}QD_{3x(n)} - \hat{C}QD_{1x(n)}} \cdot \frac{(\hat{C}QD_{3x(n)}) + \hat{C}QD_{1x(n)}}{\hat{C}QD_{3x(n)} - \hat{C}QD_{1x(n)}}
\]

Similarly, \( \hat{R}_2 \) and \( \hat{R}_3 \) are obtained as

\[
\hat{R}_2 = \left\{1 + \frac{Q_3(x)\xi_4 - Q_1(x)\xi_3}{(Q_3(x) - Q_1(x))}\right\} \cdot \left\{1 - \frac{Q_3(x)\xi_4 + Q_1(x)\xi_3}{(Q_3(x) + Q_1(x))}\right\} - 1 \cdot \left\{1 + \frac{Q_3(x)\xi_6 + Q_1(x)\xi_5}{(Q_3(x) - Q_1(x))}\right\}
\]

\[
\hat{R}_3 = \left\{1 + \frac{Q_3(x)\xi_4 - Q_1(x)\xi_3}{(Q_3(x) - Q_1(x))}\right\} \cdot \left\{1 - \frac{3Q_3(x)\xi_4 + Q_1(x)\xi_3}{(Q_3(x) + Q_1(x))}\right\} - 1 \cdot \left\{1 + \frac{3Q_3(x)\xi_6 + Q_1(x)\xi_5}{(Q_3(x) + Q_1(x))}\right\}
\]

The mean squared error of \( \hat{R}_c \) is given by as

\[
MSE(\hat{R}_c) = CQD^2 \cdot E\left\{\frac{2Q_3(x)Q_3(y)}{Q_3(x) - Q_1(y)}(\xi_2 - \xi_1) - \frac{2Q_1(x)Q_3(x)}{Q_3(x) - Q_1(x)} \ln(\xi_6 - \xi_5 + \xi_4 - \xi_3) + \cdots\right\}^2
\]

where \( \ln = c_1 + 2c_2 + 3c_3 \).

Bias and MSE of \( \hat{R}_c \) is given by

\[
\text{Bias}(\hat{R}_c) = CQV\left\{\frac{\ln}{n} \cdot \frac{1}{f(Q_3(y))^2} + \frac{1}{f(Q_1(y))^2} + \frac{2}{3}\left\{\frac{f(Q_3(y))^2}{f(Q_1(y))f(Q_3(y))}\right\}
\right\}
\]

\[
- \left\{\frac{1}{f(Q_3(y))^2} + \frac{1}{f(Q_1(y))^2}\right\} - \ln\left\{\frac{1}{f(Q_3(y))^2} + \frac{1}{f(Q_1(y))^2}\right\}
\]

Sohail et al. (2021)
Almost Unbiased Estimation of Coefficient of Dispersion from Incomplete Data

\[
\begin{align*}
&\quad + \left(\frac{2}{3}\right) \frac{1}{\left[f(Q_{1(x)}f(Q_{3(x)})\right]} - \left\{\frac{1}{(Q_{3(x)} - Q_{1(x)})^2} + \frac{1}{(Q_{3(x)} + Q_{1(x)})^2}\right\} \\
\text{and} \\
MSE(\tilde{\gamma}_c) &\cong CQD^2 \left\{ \left(\frac{2Q_1(y)Q_3(y)}{(Q_3(y) - Q_1(y))} \right)^2 (\zeta_2 - \zeta_1)^2 + \left(\frac{2Q_1(x)Q_3(x)}{(Q_3(x) - Q_1(x))} \right)^2 \kappa^2 \\
&\quad (\zeta_6 - \zeta_5 + \zeta_4 - \zeta_3)^2 - 2 \left(\frac{2Q_1(y)Q_3(y)}{(Q_3(y) - Q_1(y))} \right) \left(\frac{2Q_1(x)Q_3(x)}{(Q_3(x) - Q_1(x))} \right) \kappa \\
&\quad (\zeta_2 - \zeta_1) (\zeta_6 - \zeta_5 + \zeta_4 - \zeta_3) \right\} \\
\cong CQD^2 &\left\{ \left(\frac{2Q_1(y)Q_3(y)}{(Q_3(y) - Q_1(y))} \right)^2 \Pi(n_1,n) \left(\frac{3}{16}\right) \left\{\frac{1}{Q_3(y)^2} \left(\frac{1}{f(Q_3(y))} \right)^2 \\
&\quad + \frac{1}{Q_1(y)^2} \left(\frac{1}{f(Q_1(y))} \right)^2 - \left(\frac{2}{3}\right) \frac{1}{Q_1(y)^2} \left(\frac{1}{f(Q_1(y))} \right)^2 \\
&\quad + \frac{1}{Q_3(y)^2} \left(\frac{1}{f(Q_3(y))} \right)^2 \right\} \right\} \\
&\quad - \left\{ \left(\frac{2Q_3(y)Q_3(x)}{(Q_3(x) - Q_1(x))} \right)^{1/2} \left(\frac{P_{Q_3(y)Q_3(x)}}{(Q_3(y) - Q_1(y))} \right) - \left(\frac{3}{16}\right) \frac{1}{Q_3(y)^2} \left(\frac{1}{f(Q_3(y))} \right)^2 \\
&\quad - \left(\frac{3}{16}\right) \frac{1}{Q_1(y)^2} \left(\frac{1}{f(Q_1(y))} \right)^2 \right\} \right\} \\
&\quad + \left\{ \left(\frac{2Q_3(y)Q_3(x)}{(Q_3(x) - Q_1(x))} \right)^{1/2} \left(\frac{P_{Q_3(y)Q_3(x)}}{(Q_3(y) - Q_1(y))} \right) - \left(\frac{3}{16}\right) \frac{1}{Q_3(y)^2} \left(\frac{1}{f(Q_3(y))} \right)^2 \\
&\quad + \left(\frac{3}{16}\right) \frac{1}{Q_1(y)^2} \left(\frac{1}{f(Q_1(y))} \right)^2 \right\} \right\} \\
\text{The optimum value of } \kappa \text{ is obtained by minimizing (4.8), so we have} \\
\kappa_c = \frac{\left\{ \left(\frac{2Q_3(y)Q_3(x)}{(Q_3(x) - Q_1(x))} \right)^{3/2} \left(\frac{1}{Q_3(y)^2} \left(\frac{1}{f(Q_3(y))} \right)^2 + \frac{1}{Q_1(y)^2} \left(\frac{1}{f(Q_1(y))} \right)^2 - \frac{2}{3} \right) \left(\frac{1}{Q_1(y)^2} \left(\frac{1}{f(Q_1(y))} \right)^2 \right) \right\}}{\left\{ \left(\frac{2Q_1(y)Q_3(x)}{(Q_3(x) - Q_1(x))} \right)^{3/2} \left(\frac{1}{Q_3(y)^2} \left(\frac{1}{f(Q_3(y))} \right)^2 + \frac{1}{Q_1(y)^2} \left(\frac{1}{f(Q_1(y))} \right)^2 - \frac{2}{3} \right) \left(\frac{1}{Q_1(y)^2} \left(\frac{1}{f(Q_1(y))} \right)^2 \right) \right\}} \\
\text{The minimum mean square error of } MSE(\tilde{\gamma}_c) \text{ is} \\
\text{MSE}(\tilde{\gamma}_c) &\cong CQD^2 \left\{ \left(\frac{2Q_3(y)Q_3(x)}{(Q_3(y) - Q_1(y))} \right)^2 \Pi(n_1,n) \left(\frac{3}{16}\right) \left\{\frac{1}{Q_3(y)^2} \left(\frac{1}{f(Q_3(y))} \right)^2 \\
&\quad + \frac{1}{Q_1(y)^2} \left(\frac{1}{f(Q_1(y))} \right)^2 - \left(\frac{2}{3}\right) \frac{1}{Q_1(y)^2} \left(\frac{1}{f(Q_1(y))} \right)^2 \right\} \right\} \\
&\quad - \left\{ \left(\frac{2Q_3(y)Q_3(x)}{(Q_3(x) - Q_1(x))} \right)^{1/2} \left(\frac{P_{Q_3(y)Q_3(x)}}{(Q_3(y) - Q_1(y))} \right) - \left(\frac{3}{16}\right) \frac{1}{Q_3(y)^2} \left(\frac{1}{f(Q_3(y))} \right)^2 \\
&\quad - \left(\frac{3}{16}\right) \frac{1}{Q_1(y)^2} \left(\frac{1}{f(Q_1(y))} \right)^2 \right\} \right\} \\
&\quad + \left\{ \left(\frac{2Q_3(y)Q_3(x)}{(Q_3(x) - Q_1(x))} \right)^{1/2} \left(\frac{P_{Q_3(y)Q_3(x)}}{(Q_3(y) - Q_1(y))} \right) - \left(\frac{3}{16}\right) \frac{1}{Q_3(y)^2} \left(\frac{1}{f(Q_3(y))} \right)^2 \\
&\quad + \left(\frac{3}{16}\right) \frac{1}{Q_1(y)^2} \left(\frac{1}{f(Q_1(y))} \right)^2 \right\} \right\} \right\}
\end{align*}
\]
\[
\begin{align*}
\{(P_{Q_3(y)Q_3(x)} - \frac{9}{16})(Q_{3(y)Q_3(x)}(f(Q_{3(y)})f(Q_{3(x)}))
& - (P_{Q_3(y)Q_1(x)} - \frac{3}{16})(Q_{3(y)Q_1(x)}(f(Q_{3(y)})f(Q_{1(x)})))
& - (P_{Q_1(y)Q_3(x)} - \frac{3}{16})(Q_{1(y)Q_3(x)}(f(Q_{1(y)})f(Q_{3(x)})))
& + (P_{Q_1(y)Q_1(x)} - \frac{1}{16})(Q_{1(y)Q_1(x)}(f(Q_{1(y)})f(Q_{1(x)})))\}\]
\]
which is exactly equal to the mean square error of the linear regression estimator.

**Preparation of Antibiotic for Ratio Type Estimators**

From (8), (9), and (10), we have to prepared an antibiotic to cure ratio estimator from bias on the line of Singh and Biradar (1992) is consisting of following equations, as
\[
\sum_{i=1}^{3} \mathcal{R}_i = 1
\]

\[
\sum_{i=1}^{3} i \mathcal{R}_i = \mathbb{k}
\]

From (17) and (18), we have three unknowns to be determined from only one equation, which is not possible to find out the unique amount of antibiotic \(c_i\)s \((i = 1, 2, 3)\), to used.

In order to get the unique value of antibiotic, we should connect the antibiotic with drip of glucose using linear constrains.
\[
\sum_{i=1}^{3} c_i B(\mathcal{R}_i) = 0
\]
where \(B(\mathcal{R}_i)\) denotes the amount of bias in the \(i^{th}\) estimator of \(\mathcal{R}_i\). Using the (8), (17) and (18) we have
\[
\begin{bmatrix}
1 & 1 & 1 \\
B(\mathcal{R}_1) & B(\mathcal{R}_2) & B(\mathcal{R}_3)
\end{bmatrix}_{3 \times 3}
\begin{bmatrix}
c_1 \\
c_2 \\
c_3
\end{bmatrix}_{3 \times 1}
= \begin{bmatrix}
1 \\
\mathbb{k}
\end{bmatrix}_{3 \times 1}
\]

Bias of the \(\mathcal{R}_i\) estimator is reported up to first order approximation
\[
B(\mathcal{R}_i) \approx CQD \prod_{(n,m)}^\mathbb{k} \left(\frac{2Q_{1(x)}Q_{3(x)}}{(Q_{1(x)}-Q_{1(y)})^2} - \frac{3}{2} \frac{(Q_{1(x)}-Q_{3(y)})^2}{2f(Q_{3(x)})^2} \right) + 
\frac{1}{16}(Q_{1(y)}Q_{3(y)}(f(Q_{1(y)})f(Q_{3(y)}))) - (P_{Q_3(y)Q_1(x)} - \frac{3}{16})
\]
putting (21) in (20), we can obtain the amount of antibiotic up to first order of approximation, as
Almost Unbiased Estimation of Coefficient of Dispersion from Incomplete Data

\begin{align*}
  c_1 &= 3 - 3k + k^2 \\
  c_2 &= -3 + 5k - 2k^2 \\
  c_3 &= 1 - 2k + k^2
\end{align*}

(22) (23) (24)

This is the amount of antibiotic, which is be helpful to heal an ratio estimator from microbe of bias.

Numerical Illustration

In this section, we consider a simulated data set on the line of Singh and Horn (2000) in order to obtain the relative loss due to imputation of the proposed unbiased ratio estimator over the existing naive estimator. A hypothetical population is generated for the relative comparison on the line of Mohamed et al. (2018).

\begin{align*}
  y_j &= \mu_y + \sqrt{(1 - \rho_{yx}^2)}y_j^* + \rho_{yx}\frac{S_y}{S_x}x_j^* \text{ and } x_j = \mu_x + x_j^*
\end{align*}

(25)

where \( y_j^* \) is generated form gamma distribution with parameter \( a_y \) and \( b_y \) are the shape and scale parameter respectively, then, the study variable have the men value \( a_yb_y \) and variance \( a_yb_y^2 \). The auxiliary variable \( (x_j^*) \) is also generated from gamma distribution with mean \( a_xb_x \) and variance \( a_xb_x^2 \). The percentage relative efficiency of the unbiased ration estimator estimator \( \mathcal{R}_c \) over naive estimator is given by

\begin{align*}
  PRE(\cdot) &= \frac{V(C_{QD_{naive}e})}{MSE(\mathcal{R}_c)}
\end{align*}

(26)

For the different parametric values of the gamma distribution the percentage relative efficiency (PRE) of the reported in Table 1 and in Figure 1. The values of scale parameters are select randomly as \( b_y = 2 \) and \( b_x = 2.5 \), because its not effect the performance of the estimation procedure.

| \( a_y \) | \( a_x \) | \( \rho_{yx} \) | PRE(\cdot) | \( a_y \) | \( a_x \) | \( \rho_{yx} \) | PRE(\cdot) |
|---|---|---|---|---|---|---|---|
| .50 | 1.50 | .60 | 126.30 | 2.50 | 2.50 | .80 | 161.20 |
| .50 | 1.50 | .70 | 137.50 | 2.50 | 2.50 | .90 | 203.50 |
| .50 | 1.50 | .80 | 156.20 | 2.50 | 3.00 | .60 | 120.10 |
| .50 | 1.50 | .90 | 197.70 | 2.50 | 3.00 | .70 | 132.50 |
| .50 | 2.00 | .60 | 121.00 | 2.50 | 3.00 | .80 | 148.70 |
| .50 | 2.00 | .70 | 130.90 | 2.50 | 3.00 | .70 | 132.50 |
| .50 | 2.00 | .80 | 148.00 | 2.50 | 3.00 | .70 | 132.50 |
| .50 | 2.00 | .90 | 174.90 | 2.50 | 3.00 | .70 | 132.50 |
| .50 | 2.50 | .60 | 116.40 | 2.50 | 3.50 | .80 | 148.30 |
| .50 | 2.50 | .70 | 123.40 | 2.50 | 3.50 | .90 | 177.60 |
| .50 | 2.50 | .80 | 135.10 | 2.50 | 3.50 | .90 | 177.60 |
| .50 | 2.50 | .90 | 156.00 | 2.50 | 3.50 | .90 | 177.60 |
| .50 | 3.00 | .60 | 116.00 | 3.00 | 1.50 | .80 | 177.10 |
| .50 | 3.00 | .70 | 122.40 | 3.00 | 1.50 | .90 | 249.50 |
| .50 | 3.00 | 0.80 | 134.30 | 3.00 | 2.00 | 0.60 | 128.50 |
| .50 | 3.00 | 0.90 | 151.30 | 3.00 | 2.00 | 0.70 | 144.30 |
| .50 | 3.50 | 0.60 | 115.60 | 3.00 | 2.00 | 0.80 | 172.30 |
| .50 | 3.50 | 0.70 | 121.10 | 3.00 | 2.00 | 0.90 | 239.00 |
| .50 | 3.50 | 0.80 | 130.10 | 3.00 | 2.50 | 0.60 | 129.00 |
| .50 | 3.50 | 0.90 | 145.00 | 3.00 | 2.50 | 0.70 | 143.50 |
| .00 | 1.50 | 0.60 | 131.60 | 3.00 | 2.50 | 0.80 | 170.10 |
| .00 | 1.50 | 0.70 | 145.90 | 3.00 | 3.00 | 0.60 | 120.10 |
| .00 | 1.50 | 0.90 | 230.00 | 3.00 | 3.00 | 0.70 | 194.40 |
| .00 | 2.00 | 0.60 | 126.40 | 3.00 | 3.00 | 0.80 | 151.10 |
| .00 | 2.00 | 0.70 | 139.30 | 3.00 | 3.00 | 0.90 | 194.40 |
| .00 | 2.00 | 0.80 | 158.60 | 3.00 | 3.50 | 0.60 | 119.50 |
| .00 | 2.00 | 0.90 | 204.90 | 3.00 | 3.50 | 0.70 | 131.90 |
| .00 | 2.50 | 0.60 | 121.00 | 3.00 | 3.50 | 0.80 | 150.70 |
| .00 | 2.50 | 0.70 | 132.10 | 3.00 | 3.50 | 0.90 | 182.80 |
| .00 | 2.50 | 0.80 | 152.20 | 3.50 | 1.50 | 0.60 | 125.00 |
| .00 | 2.50 | 0.90 | 184.90 | 3.50 | 1.50 | 0.70 | 137.20 |
| .00 | 3.00 | 0.60 | 119.70 | 3.50 | 1.50 | 0.80 | 159.70 |
| .00 | 3.00 | 0.70 | 130.10 | 3.50 | 1.50 | 0.90 | 228.30 |
| .00 | 3.00 | 0.80 | 144.50 | 3.50 | 2.00 | 0.60 | 128.80 |
| .00 | 3.00 | 0.90 | 171.80 | 3.50 | 2.00 | 0.70 | 142.20 |
| .00 | 3.50 | 0.60 | 116.20 | 3.50 | 2.00 | 0.80 | 170.30 |
| .00 | 3.50 | 0.70 | 124.80 | 3.50 | 2.00 | 0.90 | 238.80 |
| .00 | 3.50 | 0.80 | 138.10 | 3.50 | 2.50 | 0.60 | 124.10 |
| .00 | 3.50 | 0.90 | 158.50 | 3.50 | 2.50 | 0.70 | 141.20 |
| .50 | 1.50 | 0.60 | 129.50 | 3.50 | 2.50 | 0.80 | 165.60 |
| .50 | 1.50 | 0.70 | 145.60 | 3.50 | 2.50 | 0.90 | 227.60 |
| .50 | 1.50 | 0.80 | 171.30 | 3.50 | 3.00 | 0.60 | 124.70 |
| .50 | 1.50 | 0.90 | 229.70 | 3.50 | 3.00 | 0.70 | 138.00 |
| .50 | 2.00 | 0.60 | 128.90 | 3.50 | 3.00 | 0.80 | 160.40 |
| .50 | 2.00 | 0.70 | 143.00 | 3.50 | 3.00 | 0.90 | 211.60 |
| .50 | 2.00 | 0.80 | 168.60 | 3.50 | 3.50 | 0.60 | 121.90 |
| .50 | 2.00 | 0.90 | 230.80 | 3.50 | 3.50 | 0.70 | 136.10 |
| .50 | 2.50 | 0.60 | 125.30 | 3.50 | 3.50 | 0.80 | 158.30 |
| .50 | 2.50 | 0.70 | 139.20 | 3.50 | 3.50 | 0.90 | 206.40 |
Almost Unbiased Estimation of Coefficient of Dispersion from Incomplete Data

Figure 1: PRE of the proposed unbiased ratio estimators over the naïve estimator against the various values of $a_y$, $a_x$ and $\rho_{yx}$

For the closer look to numerical results, we display our findings in Figure 1, as, for the different values of $a_y$ and $a_x$ over the range of 1.5 to 3.5 with an interval 0.5. Overall, our proposed almost unbiased ratio estimator remains more efficient as compare to the traditional naive estimator. The proposed procedure is relatively more efficient as the value of $a_y$ and $a_x$ increased. Thus, in the simulation study we also change the value of correlation coefficient $\rho_{yx}$ between 0.6 to 0.9. From (26), the value of PRE is seems to be free from the correlation relation coefficient. For the values of $a_y = 1.5$ and $a_x = 1.5$, the values of PRE are varies from 124.30 to 193.60 over the different values of $\rho_{yx} = 0.6$ to 0.9.

Conclusions
The efficient use of the auxiliary information is helpful to improve the efficiency of the estimation procedure over the naïve estimator consider by Bonett (2006). The proposed linearized version of estimation procedure for coefficient of dispersion is provide the unbiased ratio estimator in two phase sampling for the by imputing the missing responses. We strongly recommend to extended for the case of missing scrambled responses from the equal and un-equal clustered size populations.

Aknowledgement
Authors are thankful to the editor and reviewers for their valueable comments and suggestions.
References
Ambati, S. K., Sedory, S. A., and Singh, S. (2018). Estimation of coefficient of dispersion using auxiliary information. Communications in Statistics-Simulation and Computation, 47(7):1996–2009.
Bonett, D. G. (2006). Confidence interval for a coefficient of quartile variation. Computational Statistics & Data Analysis, 50(11):2953–2957.
Cochran, W. (1940). The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. The Journal of Agricultural Science, 30(02):262–275.
Hansen, M. H. and Hurwitz, W. N. (1946). The problem of non-response in sample surveys. Journal of the American Statistical Association, 41(236):517–529.
Heitjan, D. F. and Basu, S. (1996). Distinguishing “missing at random” and “missing completely at random”. The American Statistician, 50(3):207–213.
Little, R. and Rubin, D. B. (1987). Statistical analysis with missing data. New York: Wiley.
Longford, N. T. (2006). Missing data and small-area estimation: Modern analytical equipment for the survey statistician. Springer Science & Business Media.
Mohamed, C., Sedory, S. A., and Singh, S. (2018). A fresh imputing survey methodology using sensible constraints on study and auxiliary variables: dubious random non-response. Journal of Statistical Computation and Simulation, 88(7):1273–1294.
Rubin, D. B. (1976). Inference and missing data. Biometrika, 63(3):581–592.
Rubin, D. B. (1978). Multiple imputations in sample surveys—a phenomenological Bayesian approach to nonresponse. American Statistical Association, 1:20–34.
Rubin, D. B. (2004). Multiple imputation for nonresponse in surveys, volume 81. John Wiley & Sons.
Sarndal, C.-E. (1992). Methods for estimating the precision of survey estimates when imputation has been used. Survey methodology, 18(2):241–252.
Singh, H., Katyar, N., and Gangwar, D. (1996). A class of almost unbiased regression-type estimators in two-phase sampling applying quenouille’s method. J. Indian Soc. Agri. Stat, 48(1):98–104.
Singh, H. P. and Biradar, R. (1992). Almost unbiased ratio-cum-product estimators for the finite population mean. Test, 1(1):19–29. Singh, S. and Deo, B. (2003). Imputation by power transformation. Statistical Papers, 44(4):555–579.
Singh, S. and Horn, S. (2000). Compromised imputation in survey sampling. Metrika, 51(3):267–276.
Singh, S. and Singh, R. (1993). A new method: Almost separation of bias precipitates in sample surveys. Journal of Indian Statistical Association, 31:99–105.