Generalized Family of Estimators for Imputing Scrambled Responses

Muhammad U. Sohail, Javid Shabbir, Cem Kadilar and Shakeel Ahmed

1Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan.
2Department of Statistics, Hacettepe University, Beytepe, Ankara, Turkey.

Received: 19/07/2017, Revision received: 22/01/2018, Published online: 05/07/2018

Abstract. When there is a high correlation between the study and auxiliary variables, the rank of the auxiliary variable also correlates with the study variable. Then, the use of the rank as an additional auxiliary variable may be helpful to increase the efficiency of the estimator of the mean or the total number of the population. In the present study, we propose two generalized families of estimators for imputing the scrambling responses by using the variance and the rank of the auxiliary variable. Expressions for the bias and the mean squared error are obtained up to the first order of approximation. A numerical study is carried out to observe the performance of estimators.

Keywords. Higher order moments, Imputation, Missing data, Scrambled responses, Sensitive attribute.

MSC: 62D05.

Corresponding Author: Muhammad Umair Sohail (umairsohailch@gmail.com)
Javid Shabbir (javidshabbir@gmail.com)
Cem Kadilar (cemkadilar@gmail.com)
Shakeel Ahmed (shakeelatish05@gmail.com)
1 Introduction

In most of socio-economic surveys, missing value is a common problem that happens for many reasons, such as asking sensitive or embarrassing questions and unavailability of the respondent. Situations like asking a female ‘how many times have you induced abortions?’, questions related to income, tax evasion, etc., can be considered as sensitive questions. Initially, for these situations, Warner (1965) introduced the randomized response technique (RRT) to handle the problem of non-response relevant to the sensitive issues of the society. A randomized device was used to collect the responses from the respondents and protected their privacy despite of getting direct response. After Warner, various randomized response models were used for estimating the proportion of the nominal variable in the population (see e.g., Greenberg et al. (1969), Folsom et al. (1973), Mangat and Singh (1990), Gupta et al. (2010)).

Greenberg et al. (1971), Eichhorn and Hayre (1983), Mangat and Singh (1990), Gupta et al. (2013) considered the simple and multistage scrambling models (MSM) for estimating the mean of sensitive variables. Chaudhuri and Adhikary (1990) introduced a scrambling response model, where the j-th respondent in the sample should select two cards randomly and independently. The card $S_{1j}$ is selected from a box containing $m$ cards with a known mean ($\theta_1$) and a known variance ($\sigma_1^2$). The second card $S_{2j}$ is selected from $t$ cards with a known mean ($\theta_2$) and a known variance ($\sigma_2^2$). Let $y_j$ be the actual status of the j-th respondent in the sample (s). The j-th respondent reports the scrambled value/response $z_j$ as

$$z_j = S_{1j}y_j + S_{2j}. \quad (1.1)$$

In spite of using RRT, missing observations can occur due to reasons aside from sensitive issues. Then, how can one handle such incomplete data sets? In such circumstances, imputation is one of the most reliable methods to sort out such a problem and to build reliable data sets for a valid inference about the stated population. Many imputation techniques have been developed and considered the problem in different ways (see e.g., Rubin (1976), Heitjan and Basu (1996), Ahmed et al. (2006), Singh et al. (2010), Mohamed et al. (2017)). In Figure 1, we illustrate the imputation procedure for imputing the missing values.

Mohamed et al. (2017) considers an RRT, in which the j-th respondent is asked to select two cards (say $S_1$ and $S_2$) from two decks of cards (say $\Delta_1$ and $\Delta_2$), respectively.
The $j$-th respondent can report the scrambled response as
\[ z_j = \frac{S_1 y_j + S_2 - \theta_2}{\theta_1}. \] (1.2)

Let $E_1$ and $V_1$ be the expected value and variance with respect to the randomization device, respectively. We remark that all $E_1(S_1) = \theta_1$, $E_1(S_2) = \theta_2$, $V_1(S_1) = \sigma_1^2$, and $V_1(S_2) = \sigma_2^2$ are known, and we have $V_1(z_j) = \frac{\sigma_1^2 y_j^2 + \sigma_2^2}{\theta_1^2} = C_{\theta_1}^2 y_j^2 + (\frac{\theta_2}{\theta_1})^2 C_{\theta_2}^2$, where $C_{\theta_1}^2 = (\frac{\sigma_y}{\theta_1})^2$ and $C_{\theta_2}^2 = (\frac{\sigma_y}{\theta_2})^2$.

Let $s$ be a simple random sample of size $n$ drawn from the population $(\Omega)$ having $N$ units, and let $r$ be the total number of respondents in the subset $A$ of the sample $s$ that belongs to the sensitive characteristics with the help of the mentioned randomization scheme. Here, $(n - r)$ are those who belong to $A'$, the subset of $s$, who refuse to answer the question, therefore, $s = A \cup A'$. It is also known that $\bar{z}_r = \frac{1}{r} \sum_{j=1}^{r} z_j$ is the sample mean of the scrambling response obtained from the group $A$. Then, we have the lemma 1.1 as follows.

**Lemma 1.1.** The variance of $\bar{z}_r$ is given by
\[ V(\bar{z}_r) \approx \left(1 - \frac{1}{rN}\right)S_y^2 + \frac{1}{r} C_{\theta_1}^2 \left(1 + \frac{(N-1)C_y^2}{N}\right) C_{\theta_2}^2 + \left(\frac{\theta_2}{\theta_1}\right)^2 C_{\theta_2}^2, \] (1.3)
where $S_y^2$, $C_y^2$, $C_{\theta_1}^2$ and other parameters are defined in Appendix.

**Proof.** See Mohamed et al. (2017) for proof. □

The rest of the article is outlined as follows. Two generalized families of estimators for imputing scrambling response are proposed by using higher order moments along with the rank of the auxiliary variable. Theoretical comparison of the proposed generalized imputation methods over the mean method of imputation is considered in Section 3. For evaluating the relative performance of the proposed generalized families, a numerical study is conducted for various choices of the constants in Section 4. Finally, Section 5 concludes the article.
2 Proposed Estimators

In survey sampling, it is a common practice to incorporate known auxiliary information in the estimation stage of population characteristics for improving efficiency. The use of the auxiliary information can increase the precision of the estimators both at estimation and design stages. The traditional ratio, product and classical regression estimators are used, when there is a high correlation between the study and the auxiliary variables. If the correlation between the study variable and the auxiliary variable is sufficiently large, then the rank of the auxiliary variable is also correlated with the study variable. The inclusion of the rank of a variable may help to improve the efficiency of the estimators. Let \( \bar{y}, \bar{x} \) and \( \bar{r} \) be the sample means and \( \bar{Y}, \bar{X} \) and \( \bar{R} \) be the population means of the study variable, the auxiliary variable, and the rank of the auxiliary variable, respectively. Let \( s_y^2 = \sum_{j=1}^{n} (y_j - \bar{y})^2 / (n - 1) \), \( s_x^2 = \sum_{j=1}^{n} (x_j - \bar{x})^2 / (n - 1) \) and \( s_r^2 = \sum_{j=1}^{n} (r_j - \bar{r})^2 / (n - 1) \) be the unbiased sample variances corresponding to the population variances \( S_y^2 = \sum_{j=1}^{N} (y_j - \bar{Y})^2 / (N - 1) \), \( S_x^2 = \sum_{j=1}^{N} (x_j - \bar{X})^2 / (N - 1) \), and \( S_r^2 = \sum_{j=1}^{N} (r_j - \bar{R})^2 / (N - 1) \) of \( y, x, \) and \( r \), respectively. We propose two generalized families of estimators for the imputation by using higher order moments and rank of the auxiliary variable. We consider ratio and regression types of estimators in the following subsections.

2.1 Generalized Ratio Type Estimators

The generalized ratio method of imputation for imputing scrambling response is

\[
\hat{y}_{s,CR} = \frac{\sum_{i = 1}^{N} x_i}{\sum_{i = 1}^{N} z_i} \left( \frac{n z}{n - 1} \right) \left( \frac{s_x^2}{s_r^2} \right) \left( \frac{s_y^2}{s_r^2} \right) \left( \frac{\bar{x}}{\bar{r}} \right) \left( \frac{\bar{x}}{\bar{r}} \right) - \frac{r^2}{(n - 1)} \right.
\]

where \( g_1, g_2 \) and \( g_3 \) are unknown constants and their values are to be determined by minimizing the mean squared error and \( (\bar{x}_n \and \bar{x}_r), (s_x^2 \and s_r^2) \), and \( (\bar{r}_n \and \bar{r}_r) \) are
the mean, variance and rank of the auxiliary variable for \( n \) and \( r \) units, respectively. We define the \( \bar{y}_{sGR} \) in the form of error terms given in the Appendix as

\[
\bar{y}_{sGR} = \bar{Y}(1 + e_0)(1 + e_1)^s_1(1 + e_1)^{-s_1}(1 + e_3)^s_2(1 + e_3)^{-s_2}(1 + e_6)^{-s_3}
\]

\[
= \bar{Y}\left[\frac{g_1(g_1 + 1)}{2} e_1^2 + \frac{g_1(g_1 - 1)}{2} e_2^2 + \frac{g_3(g_3 + 1)}{2} + \frac{g_3(g_3 - 1)}{2} e_3^2\right]
\]

\[
+ \frac{g_2(g_2 + 1)}{2} e_4^2 - (g_1 e_0 e_1 + g_2 e_0 e_3 + g_3 e_0 e_6)
\]

\[
+ g_1 g_2 e_1 e_3 + g_1 g_2 e_1 e_6 + g_2 g_3 e_3 e_6 - (g_1 e_2^2 + g_4 e_4^2 + g_5 e_5^2).
\]  \tag{2.3}

The bias and mean squared error of \( \bar{y}_{sGR} \) are given by

\[
\text{Bias}(\bar{y}_{sGR}) = \left[\frac{\theta_{rN}}{2} + \frac{\theta_{nN}}{2}\right] \bar{Y} C_x^2 + \left[\frac{\theta_{rN}}{2} + \frac{\theta_{nN}}{2}\right] \bar{Y} C_r^2
\]

\[
+ \left[\frac{\theta_{rN}}{2} + \frac{\theta_{nN}}{2}\right] \bar{Y} (\lambda_{040} - 1)
\]

\[- \theta_{rN} \left[\frac{1}{2} g_1 \rho_{xy} C_y C_x + g_2 \rho_{xy} y C_x + g_3 \rho_{xy} y C_r + g_3 \rho_{x} C_y C_r + g_4 g_5 \rho_{xy} C_r \lambda_{030}\right]
\]

\[- \theta_{nN} \left[\frac{1}{2} g_1 C_x^2 + g_3 (\lambda_{040} - 1) + g_3 C_r^2\right].
\]  \tag{2.4}

and

\[
\text{MSE}(\bar{y}_{sGR}) \approx \theta_{nN} S_y^2 + \frac{1}{r} \left[\bar{Y}^2 C_{\theta_1} \left\{1 + \frac{(N - 1) C_y^2}{N}\right\} + \frac{\theta_{rN}^2}{\theta_{nN}^2} C_{\theta_2}^2\right] + \theta_{rN} S_y^2
\]

\[
+ g_1^2 S_x^2 + g_2^2 S_x^2 (\lambda_{040} - 1) + g_3^2 S_r^2 - 2 \left(g_1 S_{xy} + g_2 S_x^2 S_y \lambda_{120}\right)
\]

\[
+ g_1 S_{xy} - g_1 g_2 S_x^3 \lambda_{030} - g_2 g_3 S_{xy} - g_3 g_3 S_r S_y \lambda_{003}\right]\]  \tag{2.5}

The optimum values of \( g_1, g_2, \) and \( g_3 \) are obtained respectively as

\[
g_{1(\text{opt.})} = \frac{w_1}{t_1} = G_1, \quad g_{2(\text{opt.})} = \frac{w_2}{t_2} = G_2, \quad \text{and} \quad g_{3(\text{opt.})} = \frac{w_3}{t_3} = G_3.
\]  \tag{2.6}
where

\[
\begin{align*}
    w_2 &= S_x^2 S_y^2 \lambda_{120} (S_x^2 S_r^2 - S_x^2) - S_{xy} S_x \lambda_{030} (S_x^2 S_r^2 - S_x^2) - (S_{ry} S_x^2 - S_{xy} S_{rx}) \\
         &\quad (S_x^2 S_r^2 \lambda_{003} - S_x S_{xr} \lambda_{030}), \\
    t_2 &= (S_x^2 S_r^2 - S_x^2) (\lambda_{040} - 1) S_x^4 - (S_r S_x^2 \lambda_{003} - S_{rx} S_r S_x^3 \lambda_{030}), \\
    w_3 &= S_{ry} S_x^2 t_2 - S_{xr} S_{xy} t_2 - w_2 (S_r S_x^2 \lambda_{003} - S_{rx} S_x S_r \lambda_{030}), \\
    t_3 &= t_2 (S_x^2 S_r^2 - S_x^2), \\
    w_1 &= S_{xy} t_2 t_3 - S_{xr} t_2 w_3 - w_2 t_3 S_x^3 \lambda_{030} \text{ and } t_1 = t_2 t_3 S_x^2.
\end{align*}
\]

Substituting the optimum values of \( g_q \) for \( q = 1, 2, 3 \) in (2.5), we get the minimum mean squared error of \( \hat{y}_{SG_k} \) as

\[
MSE(\hat{y}_{SG_k})_{\text{min}} \approx \theta_{y,N} S_y^2 + \frac{1}{r} \left[ \hat{Y}^2 C_0 \left\{ 1 + \frac{(N - 1)}{N} C_0 \right\} - \frac{\theta_2^2}{\theta_1^2} \right] + \theta_{x,y} S_y^2 + G_1 S_x^2 + G_2 S_x^4 (\lambda_{040} - 1) + G_3 S_r^2 \\
- 2G_1 S_{xy} - 2G_2 S_x^2 S_y \lambda_{120} - 2G_3 S_{rx} + 2G_1 G_2 S_x^3 \lambda_{030} \\
+ 2G_1 G_3 S_{xr} + 2G_2 G_3 S_r S_x^2 \lambda_{003}.
\] (2.7)

Some members of the proposed family of ratio type estimators for different choices of \( g_q \) for \( q = 1, 2, 3 \) are given in Table 1. The biases and mean squared errors of \( \hat{y}_{sk} \) for \( k = 2, 3, \cdots, 8 \) up to the first order approximation are given in Tables 2 and 3, respectively. Similarly, mean squared error of the \( \hat{y}_{sk} \) are given in Table 3.

### 2.2 Generalized Difference Method of Imputation

The generalized regression method of imputation with known regression coefficients is given by

\[
\hat{z}_j = \begin{cases} 
    z_j 
    & \text{if } j \in A \\
    \frac{z_r + d_1 \beta_1 (x_i - \bar{x}_r) + d_2 \beta_2 \left[ \frac{n(x_i - \bar{x}_r)^2}{(n-1)} + \frac{n \sum_{j=1}^{n} (x_j - \bar{x})^2}{(n-\tau)(n-1)} \right]}{d_3 \beta_3 (r_j - \bar{r})} & \text{if } j \in A^c
\end{cases}
\] (2.8)
The point estimator of the population mean is given by

$$
\hat{y}_{SC_{Bi}} = \frac{1}{n} \sum_{j \in s} z_j
= \frac{1}{n} \left[ \sum_{j \in A} z_j + \sum_{j \in A'} \left( \frac{z_j}{n} + d_1 \beta_1 (x_i - \bar{x}) + d_2 \beta_2 \frac{n(x_i - \bar{x})^2}{n - 1} \right) \right]
$$
where \(d_1, d_2\) and \(d_3\) are the unknown constants and their values are to be determined by minimizing the mean squared error. \(\beta_1, \beta_2\) and \(\beta_3\) are the population constants which are supposed to be known in advance.

In terms of errors (see Appendix), we have

\[
\tilde{y}_{SGRe} = \hat{Y}(1 + e_0) + d_1\beta_1\bar{X}(e_2 - e_1) + d_2\beta_2S_x^2(e_4 - e_3) + d_3\beta_3\bar{R}(e_7 - e_6),
\]

and

\[
MSE(\tilde{y}_{SGRe}) \approx \theta_{nN}S_y^2 + \frac{1}{r}\left[\tilde{Y}^2C_{\theta_1}\left\{1 + \frac{(N-1)C_y^2}{N}\right\} + \frac{\theta_2^2}{\theta_1^2}C_\theta^2 + \theta_{r,n}\left(S_y^2 + d_1^2\beta_1^2S_x^2 + d_2^2\beta_2^2S_x^2(\lambda_{040} - 1) + d_3^2\beta_3^2S_r^2 - 2(d_1\beta_1S_{xy} + d_2\beta_2S_xS_y\lambda_{120} + d_3\beta_3S_{ry} - d_1d_2\beta_1\beta_2S_x^3\lambda_{030} - d_1d_3\beta_1\beta_3S_{xr} - d_2d_3\beta_2\beta_3S_y^2\lambda_{003})\right\}\right].
\]

### Table 2: Bias of \(k^{th}\)-ratio estimator.

| \(k\) | Bias(\(\tilde{y}_{s_k}\)) |
|-------|-------------------|
| 2     | \(\theta_{r,n}\tilde{Y}\left(C_x^2 - \rho_{xy}C_xC_y\right)\). |
| 3     | \(\theta_{r,n}\tilde{Y}\left(\lambda_{040} - 1 - C_y\lambda_{120}\right)\). |
| 4     | \(\theta_{r,n}\tilde{Y}\left(C_r^2 - \rho_{ry}C_rC_y\right)\). |
| 5     | \(\theta_{r,n}\tilde{Y}\left[C_x^2 + (\lambda_{040} - 1) - C_x\lambda_{030} - C_y\lambda_{120} - \rho_{xy}C_xC_y\right]\). |
| 6     | \(\theta_{r,n}\tilde{Y}\left[C_x^2 + C_r^2 - \rho_{xy}C_xC_y - \rho_{ry}C_rC_y\right]\). |
| 7     | \(\theta_{r,n}\tilde{Y}\left[C_r^2 + (\lambda_{040} - 1) - C_r\lambda_{003} - C_y\lambda_{120} - \rho_{ry}C_rC_y\right]\). |
| 8     | \(\theta_{r,n}\tilde{Y}\left[C_x^2 + (\lambda_{040} - 1) + C_r^2 - C_x\lambda_{030} - C_r\lambda_{003} - \rho_{xy}C_xC_y - C_y\lambda_{120} + \rho_{xy}C_xC_r\right]\). |
Table 3: Mean squared error of the proposed estimators.

| $k$ | $MSE(y_{sk})$ |
|-----|---------------|
| 1   | $\theta_{r,n}S^2_y + \frac{1}{\tau} \left[ \bar{Y}^2 C^2_{\theta_1} \left\{ 1 + \frac{(N-1)C^2}{N} \right\} + \frac{\theta_2^2}{\theta_1^2} C^2_{\theta_2} \right]$. |
| 2   | $\theta_{r,n}S^2_y + \frac{1}{\tau} \left[ \bar{Y}^2 C^2_{\theta_1} \left\{ 1 + \frac{(N-1)C^2}{N} \right\} + \frac{\theta_2^2}{\theta_1^2} C^2_{\theta_2} \right] + \theta_{r,n} \left( S^2_y + R^2 S^2_x - 2RS_{xy} \right)$. |
| 3   | $\theta_{r,n}S^2_y + \frac{1}{\tau} \left[ \bar{Y}^2 C^2_{\theta_1} \left\{ 1 + \frac{(N-1)C^2}{N} \right\} + \frac{\theta_2^2}{\theta_1^2} C^2_{\theta_2} \right] + \theta_{r,n} \left( S^2_y + \bar{Y}^2(\lambda_{040}-1) - 2\bar{Y}S_y\lambda_{120} \right)$. |
| 4   | $\theta_{r,n}S^2_y + \frac{1}{\tau} \left[ \bar{Y}^2 C^2_{\theta_1} \left\{ 1 + \frac{(N-1)C^2}{N} \right\} + \frac{\theta_2^2}{\theta_1^2} C^2_{\theta_2} \right] + \theta_{r,n} \left( S^2_y + R^2 S^2_x - 2R^2_{S_{xy}} \right)$. |
| 5   | $\theta_{r,n}S^2_y + \frac{1}{\tau} \left[ \bar{Y}^2 C^2_{\theta_1} \left\{ 1 + \frac{(N-1)C^2}{N} \right\} + \frac{\theta_2^2}{\theta_1^2} C^2_{\theta_2} \right] + \theta_{r,n} \left[ S^2_y + \bar{Y}^2(\lambda_{040}-1) + R^2 S^2_x - 2RS_{xy} - 2R^2_{S_{xy}} + 2R^2_{S_{xy}} \right]$ |
| 6   | $\theta_{r,n}S^2_y + \frac{1}{\tau} \left[ \bar{Y}^2 C^2_{\theta_1} \left\{ 1 + \frac{(N-1)C^2}{N} \right\} + \frac{\theta_2^2}{\theta_1^2} C^2_{\theta_2} \right] + \theta_{r,n} \left[ S^2_y + R^2 S^2_x + R^2 S^2 - 2RS_{xy} - 2R^2_{S_{xy}} \right]$ |
| 7   | $\theta_{r,n}S^2_y + \frac{1}{\tau} \left[ \bar{Y}^2 C^2_{\theta_1} \left\{ 1 + \frac{(N-1)C^2}{N} \right\} + \frac{\theta_2^2}{\theta_1^2} C^2_{\theta_2} \right] + \theta_{r,n} \left[ S^2_y + \bar{Y}^2(\lambda_{040}-1) + R^2 S^2_x - 2R^2_{S_{xy}} \right]$ |
| 8   | $\theta_{r,n}S^2_y + \frac{1}{\tau} \left[ \bar{Y}^2 C^2_{\theta_1} \left\{ 1 + \frac{(N-1)C^2}{N} \right\} + \frac{\theta_2^2}{\theta_1^2} C^2_{\theta_2} \right] + \theta_{r,n} \left[ S^2_y + R^2 S^2_x + R^2 S^2 - 2RS_{xy} - 2R^2 S_{xy} \right]$ |

The optimum values of $d_1$, $d_2$ and $d_3$ are obtained by minimizing (2.10), respectively, by

$$d_{1(\text{opt})} = \frac{b_1}{b_{11}} = F_1, \quad d_{2(\text{opt})} = \frac{b_2}{b_{22}} = F_2, \quad \text{and} \quad d_{3(\text{opt})} = \frac{b_3}{b_{33}} = F_3, \quad (2.11)$$

where

$$b_1 = S^2_x \left( \lambda_{120} S_x S_y \lambda_{030} + S_{xy}(\lambda_{003} - \lambda_{040} + 1) \right) S^2$$
-λ_{003}(S_r y S_x λ_{030} + S_{rx} S_y λ_{120}) S_r + S_{ry} S_{xr}(λ_{040} - 1),

\[ b_{11} = S_{xy}\{S_r(λ^2_{003} + λ^2_{030} - λ_{040} + 1)S_x^2 - 2S_r S_{sx} \lambda_{030} λ_{003} + S_{xr}^2(λ_{040} - 1)\}, \]

\[ b_2 = S_x(λ_{040} - 1)^2\{S_r^2 S_y(\lambda_{S_y} S_y λ_{120} - S_{ry} λ_{003}) - λ_{030}(S_r S_{xy} - S_{ry} S_{xr}) \]

\[ + S_{xr}(S_r S_{xy} λ_{003} - S_{xr} S_y λ_{120})\}, \]

\[ b_{22} = S_x^2 S_y λ_{120}\{S_r^2(-λ^2_{003} - λ^2_{030} + λ_{040} - 1) + 2S_r S_{sx} \lambda_{030} λ_{003} - S_{xr}^2(λ_{040} - 1)\}, \]

\[ b_3 = S_r^2 S_x^2 λ_{030} λ_{120} - S_r^3 S_x^2 S_{xy} λ_{030} λ_{040} + S_r^2 S_{ry} S_{y}^2 λ_{003} - S_r^2 S_{rx} S_{xy} λ_{040} - S_r^2 S_{xr} S_{xy}, \]

\[ b_{33} = S_r^2 S_{ry} S_x^2 λ_{030} + S_r^2 S_y S_{x}^2 λ_{030} - λ_{040} S_r^2 S_{ry} S_{y}^2 \]

\[ - 2S_r S_{sx} S_{xr} S_y λ_{030} λ_{003} + S_r^2 S_{xy} S_x^2 + S_{ry} S_{xr}(λ_{040} - 1). \]

Substituting the optimum values of \(d_q\) for \(q = 1, 2, 3\) in (2.10), the minimum mean squared error of \(\hat{y}_{GGR}\) is obtained by

\[
\text{MSE}(\hat{y}_{GGR})_{\text{min}} \equiv \theta_{n, N} S_y^2 + \frac{1}{r} \left[ \hat{y}^2 C^2 \sigma^2 \left( 1 + \frac{(N - 1) \lambda}{N} C^2 \right) + \frac{\theta^2}{\sigma^2} C^2 \right]
\]

\[
+ \theta_{n, m} \left[ S_y^2 + F_1 S_x^2 + F_2^2 S_t^2 (λ_{040} - 1) + F_3 S_y^2 - 2F_1 S_{xy} \right]
\]

\[
- 2F_2 S_x^2 S_y λ_{120} - 2F_3 S_y + 2F_1 F_2 S_x^3 λ_{030} + 2F_1 F_3 S_{xr}
\]

\[
+ 2F_2 F_3 S_x^2 λ_{003}. \] (2.12)

The possible members of the generalized difference family are given in Table 4. The \(\text{MSE}(\hat{y}_{sa})\) for the difference estimators is given in Table 5. The optimum values of \(\beta_i\)'s are obtained by minimizing the mean squared error equations and they are given in Table 6. The minimum MSE equations of \((\hat{y}_{sa})\) are given in Table 7. If the marginal distributions of \(x\) and rank of \(x\) are normal, then \(λ_{003}\) will be zero. Hence, under the assumption of normality, the last term in \(\text{MSE}(\hat{y}_{sa(\text{mn})})\) vanishes.

### 3 Theoretical Comparisons

In this section, we consider the theoretical comparison of the proposed ratio family of estimators with the mean method of imputation.
For the ratio estimator, we have

(i) \[ \text{MSE}(\bar{y}_{s1}) > \text{MSE}(\bar{y}_{s2}), \] if
\[ 2\rho_{xy}C_y - C_x > 0. \] (3.1)

(ii) \[ \text{MSE}(\bar{y}_{s1}) > \text{MSE}(\bar{y}_{s3}), \] if
\[ \bar{Y}^2(\lambda_{04} - 1) - 2\bar{Y}S_y\lambda_{12} > 0. \] (3.2)

Table 4: Some members of the proposed and existing regression type estimators.

| \( w \) | \( d_1 \) | \( d_2 \) | \( d_3 \) | \( \bar{y}_{s,\psi} \) | Imputation Procedure |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | \( \bar{y}_r \). | \( \hat{z}_j = \begin{cases} 
\bar{z}_j & \text{if } j \in A \\
\bar{z}_r & \text{if } j \in A' 
\end{cases} \) |
| 2 | 1 | 0 | 0 | \( \bar{z}_r + \beta_1(x_n - \bar{x}_r) \). | \( \hat{z}_j = \begin{cases} 
\bar{z}_j & \text{if } j \in A \\
\bar{z}_r + \beta_1(x_i - \bar{x}_r) & \text{if } j \in A' 
\end{cases} \) |
| 3 | 0 | 1 | 0 | \( \bar{z}_r + \beta_2(s_n^2 - s_r^2) \). | \( \hat{z}_j = \begin{cases} 
\bar{z}_j & \text{if } j \in A \\
\bar{z}_r + \beta_2 \left( \frac{n(x_j - \bar{x}_n)^2}{(n-1)} + \frac{n \sum_{i \in A}(x_i - \bar{x}_n)^2}{(n-r)(n-1)} \right) & \text{if } j \in A' 
\end{cases} \) |
| 4 | 0 | 0 | 1 | \( \bar{z}_r + \beta_3(r_n - \bar{r}_r) \). | \( \hat{z}_j = \begin{cases} 
\bar{z}_j & \text{if } j \in A \\
\bar{z}_r + \beta_3(r_j - \bar{r}_r) & \text{if } j \in A' 
\end{cases} \) |
| 5 | 1 | 1 | 0 | \( \bar{z}_r + \beta_1(x_n - \bar{x}_r) + \beta_2(s_n^2 - s_r^2) \). | \( \hat{z}_j = \begin{cases} 
\bar{z}_j & \text{if } j \in A \\
\bar{z}_r + \beta_1(x_i - \bar{x}_r) + \beta_2 \left( \frac{n(x_j - \bar{x}_n)^2}{(n-1)} + \frac{n \sum_{i \in A}(x_i - \bar{x}_n)^2}{(n-r)(n-1)} \right) & \text{if } j \in A' 
\end{cases} \) |
| 6 | 1 | 0 | 1 | \( \bar{z}_r + \beta_1(x_n - \bar{x}_r) + \beta_3(r_n - \bar{r}_r) \). | \( \hat{z}_j = \begin{cases} 
\bar{z}_j & \text{if } j \in A \\
\bar{z}_r + \beta_1(x_i - \bar{x}_r) + \beta_3(r_j - \bar{r}_r) & \text{if } j \in A' 
\end{cases} \) |
\[
\begin{align*}
7 & | 1 & 1 & 0 & z_r + \beta_3(r_n - \bar{r}) + \\
& & & & \beta_2(s_n^2 - s_r^2). \\
& | 8 & 1 & 1 & 1 & z_r + \beta_3(x_n - \bar{x}) + \\
& & & & \beta_2(s_n^2 - s_r^2) + \beta_3(r_n - r_r). \\
\end{align*}
\]

\[
\hat{z}_j = \begin{cases} 
  \frac{z_j}{\sqrt{n}} + \beta_3(r_i - \bar{r}) & \text{if } j \in A \\
  + \beta_2 \left[ \frac{n(x_j - \bar{x})^2}{(1-\alpha(n-1))} + \frac{n \sum_{i \in \Delta A} (x_j - \bar{x})^2}{(n-r)(n-1)} \right] & \text{if } j \in A' \\
  - \frac{n \sum_{i \in \Delta A} (x_j - \bar{x})^2}{(n-r)(n-1)} & \text{if } j \in A \\
  + \beta_3(r_j - \bar{r}) & \text{if } j \in A' 
\end{cases}
\]

(iii) \( \text{MSE}(\hat{y}_{S1}) > \text{MSE}(\hat{y}_{S4}) \), if

\[2r_yC_y - C_r > 0.\] \hspace{1cm} (3.3)

(iv) \( \text{MSE}(\hat{y}_{S1}) > \text{MSE}(\hat{y}_{S3}) \), if

\[\rho_{xy} + \frac{1}{C_y} \left( \lambda_{030} - \frac{C_x}{2} \left( 1 - \frac{\lambda_{040} - 1}{C_x} \right) \right) - \frac{\lambda_{120}}{C_x} > 0.\] \hspace{1cm} (3.4)

(v) \( \text{MSE}(\hat{y}_{S1}) > \text{MSE}(\hat{y}_{S6}) \), if

\[\rho_{xr} - \frac{1}{2C_xC_r} \left[ C_x^2 + C_r^2 - 2C_y \left( \rho_{xy}C_x + \rho_{ry}C_r \right) \right] > 0.\] \hspace{1cm} (3.5)

(vi) \( \text{MSE}(\hat{y}_{S1}) > \text{MSE}(\hat{y}_{S3}) \), if

\[\rho_{ry} + \frac{1}{C_y} \left[ \lambda_{003} - \frac{C_r}{2} \left( 1 - \frac{\lambda_{040} - 1}{C_r} \right) \right] - \frac{\lambda_{120}}{C_r} > 0.\] \hspace{1cm} (3.6)

(vii) \( \text{MSE}(\hat{y}_{S1}) > \text{MSE}(\hat{y}_{S8}) \), if

\[\rho_{xr} - C_y \left( \frac{C_x^3 + C_r^3}{2} - \rho_{xy} + \rho_{ry}' + \lambda_{120}' - \lambda_{030}' - \lambda_{003}' \right) > 0,\] \hspace{1cm} (3.7)
where

\[ C_x' = C_x/(C_r C_y), \quad C_r' = C_r/(C_x C_y), \quad \rho_{xy} = \rho_{xy}/(C_x C_y), \quad \rho_{wy} = \rho_{wy}/(C_x C_y), \]

\[ \lambda'_{120} = \lambda_{120}/(C_x C_r), \lambda'_{030} = \lambda_{030}/(C_x C_y), \quad \lambda'_{003} = \lambda_{003}/(C_x C_y). \]

The proposed generalized ratio family of imputing the scrambling response is more efficient than the simple mean imputation method if conditions (3.1) - (3.7) are satisfied.

Table 5: Mean squared error equation of difference estimators.

| w | \( MSE(\hat{y}_{sw}) \) |
|---|-------------------|
| 1 | \[ \theta_{rN} S_y^2 + \frac{1}{r} \left[ Y^2 C_{\theta_1}^2 \left( 1 + \frac{(N-1)C_y^2}{N} \right) + \frac{\partial^2}{\partial \theta_1^2} C_{\theta_2}^2 \right] \] |
| 2 | \[ \theta_{nN} S_y^2 + \frac{1}{r} \left[ Y^2 C_{\theta_1}^2 \left( 1 + \frac{(N-1)C_y^2}{N} \right) + \frac{\partial^2}{\partial \theta_1^2} C_{\theta_2}^2 \right] + \theta_{rN} \left[ S_y^2 + \beta_2^2 S_x^2 - 2\beta_1 S_{xy} \right] \] |
| 3 | \[ \theta_{nN} S_y^2 + \frac{1}{r} \left[ Y^2 C_{\theta_1}^2 \left( 1 + \frac{(N-1)C_y^2}{N} \right) + \frac{\partial^2}{\partial \theta_1^2} C_{\theta_2}^2 \right] + \theta_{rN} \left[ S_y^2 + \beta_2^2 S_x^2 + 2\beta_2 S_x^4 (\lambda_{040} - 1) - 2\beta_2 S_y^2 \lambda_{120} \right] \] |
| 4 | \[ \theta_{nN} S_y^2 + \frac{1}{r} \left[ Y^2 C_{\theta_1}^2 \left( 1 + \frac{(N-1)C_y^2}{N} \right) + \frac{\partial^2}{\partial \theta_1^2} C_{\theta_2}^2 \right] + \theta_{rN} \left[ S_y^2 + \beta_3^2 S_y^2 - 2\beta_3 S_{xy} \right] \] |
| 5 | \[ \theta_{nN} S_y^2 + \frac{1}{r} \left[ Y^2 C_{\theta_1}^2 \left( 1 + \frac{(N-1)C_y^2}{N} \right) + \frac{\partial^2}{\partial \theta_1^2} C_{\theta_2}^2 \right] + \theta_{rN} \left[ S_y^2 + \beta_1^2 S_x^2 + \beta_3^2 S_x^2 (\lambda_{040} - 1) - 2\beta_2 S_y^2 \lambda_{120} - 2\beta_3 S_{xy} + 2\beta_1 \beta_2 S_x^2 \lambda_{030} \right] \] |
| 6 | \[ \theta_{nN} S_y^2 + \frac{1}{r} \left[ Y^2 C_{\theta_1}^2 \left( 1 + \frac{(N-1)C_y^2}{N} \right) + \frac{\partial^2}{\partial \theta_1^2} C_{\theta_2}^2 \right] + \theta_{rN} \left[ S_y^2 + \beta_1^2 S_x^2 + \beta_3 S_x^2 - 2\beta_1 S_{xy} - 2\beta_3 S_{xy} + 2\beta_1 \beta_2 S_x^2 \right] \] |
| 7 | \[ \theta_{nN} S_y^2 + \frac{1}{r} \left[ Y^2 C_{\theta_1}^2 \left( 1 + \frac{(N-1)C_y^2}{N} \right) + \frac{\partial^2}{\partial \theta_1^2} C_{\theta_2}^2 \right] + \theta_{rN} \left[ S_y^2 + \beta_2 S_x^4 (\lambda_{040} - 1) - \beta_3^2 S_y^2 - 2\beta_2 S_y^2 \lambda_{120} - 2\beta_3 S_{xy} + 2\beta_2 \beta_3 S_x^2 S_y \lambda_{003} \right] \] |
| 8 | \[ \theta_{nN} S_y^2 + \frac{1}{r} \left[ Y^2 C_{\theta_1}^2 \left( 1 + \frac{(N-1)C_y^2}{N} \right) + \frac{\partial^2}{\partial \theta_1^2} C_{\theta_2}^2 \right] + \theta_{rN} \left[ S_y^2 + \beta_1^2 S_x^2 + \beta_2 S_x^4 (\lambda_{040} - 1) + \beta_3^2 S_y^2 - 2\left( \beta_1 S_{xy} + \beta_2 S_y^2 \lambda_{120} + \beta_3 S_{xy} - \beta_1 \beta_2 S_x^2 \lambda_{030} - \beta_1 \beta_3 S_x^2 - \beta_2 \beta_3 S_x^2 \lambda_{003} \right) \right. \] \] |
4 Numerical Study

We use the data set by Murthy (967). for this data set, we have $y$ as the number of workers, and $x$ as the Capital fixed in (000) rupees. In addition, we have:

\[ \begin{align*}
N &= 80, \ n = 60, \ \bar{Y} = 1126.00, \ \bar{X} = 285.10, \ S_y = 845.61, \ S_x = 270.43, \ \lambda_{040} = 1.0892, \\
\lambda_{120} &= 2.8306, \ \lambda_{120} = 2.8306, \ \rho_{xy} = 0.9884, \ \rho_{ry} = 0.8851, \ \rho_{xr} = 0.9201. \\
\end{align*} \]

The percentage of relative efficiencies (PREs) are obtained as

\[ \begin{align*}
PRE(\cdot) &= \frac{\text{Var}(\bar{Y}_{SM})}{\text{MSE}(\bar{Y}_{SM})} \times 100, \\
PRE(k) &= \frac{\text{Var}(\bar{Y}_{SM})}{\text{MSE}(\bar{Y}_{SM})} \times 100, \\
PRE(w) &= \frac{\text{Var}(\bar{Y}_{SM})}{\text{MSE}(\bar{Y}_{SM})} \times 100, \\
\end{align*} \]

where $k, w = 2, \cdots, 8$. For the ratio of the mean of two scrambling variables and the coefficients of variation of the first and the second scrambling cards, we use

\[ \begin{align*}
\theta_2^2/\theta_1^2 &= 0.5, 1.0, 1.5 \quad \text{and} \quad C_{\theta_1}^2 = C_{\theta_2}^2 = 0.1, 0.2. \\
\end{align*} \]

Based on the mentioned data set, the PREs of the generalized estimators and their special cases are shown in Tables 8, 10, and 11 over the different values of constants at various response rates. It is clearly noticed that the proposed family outperforms according to the mean method of imputation. One more thing which we wish to remark in Tables 8, 10, and 11 is that the proposed generalized method of imputation has considerably greater efficiency at low response rates.

The summary statistics of the results are given in Table 9. Note that the maximum value of PRE(R) is 1127.1130% for the generalized ratio estimator, and for generalized regression estimator is 1069.7540%. This shows that the proposed estimators are quite efficient.

5 Conclusion

In the present study, we suggested two generalized families of estimators for imputing scrambled responses by utilizing higher order moments along with known mean of
the ranks of the auxiliary variable. Based on numerical findings, it is shown out that the proposed generalized methods are more efficient as compared to their counterpart. Thus, we recommend the use of the proposed generalized methods for efficiently estimating the finite population mean of the sensitive variable.

The current work can easily be extended to the generalized exponential, exponential-ratio and exponential-product type estimators in different sampling schemes. This article is a part of an ongoing research that will appear in forthcoming articles.

**Acknowledgements**

The authors are thankful to the reviewers for the very useful and constructive comments on the earlier version of this paper, which have led to a substantial improvement of the paper.

Table 6: Optimum values of $\beta_t$ for $t = 1, 2, \text{ and } 3$.

| $w$ | Optimum Values                                                                                               |
|-----|---------------------------------------------------------------------------------------------------------------|
| 2   | $\beta_1(\text{opt}) = \frac{S_{xy}}{S_x^2}$, $\beta_2(\text{opt}) = \frac{S_y \lambda_{120}}{S_x^2 (\lambda_{40} - 1)}$, $\beta_3(\text{opt}) = \frac{S_y}{S_x^2}$, and $\beta_{2(\text{opt})} = \frac{S_y S_y \lambda_{120} - S_{xy} \lambda_{100}}{S_x^2 (\lambda_{40} - 1) - \lambda_{100}}$, $\beta_{3(\text{opt})} = \frac{S_y}{S_x^2 (\lambda_{40} - 1)}$. |
| 3   | $\beta_1(\text{opt}) = \frac{S_{xy} (\lambda_{40} - 1) - S_x S_y \lambda_{100}}{S_x^2 (\lambda_{40} - 1) - \lambda_{100}}$, and $\beta_{2(\text{opt})} = \frac{S_y S_y \lambda_{120} - S_{xy} \lambda_{100}}{S_x^2 (\lambda_{40} - 1) - \lambda_{100}}$, $\beta_{3(\text{opt})} = \frac{S_y}{S_x^2 (\lambda_{40} - 1)}$. |
| 4   | $\beta_1(\text{opt}) = \frac{S_{xy} (\lambda_{40} - 1) - S_x S_y \lambda_{100}}{S_x^2 (\lambda_{40} - 1) - \lambda_{100}}$, and $\beta_{2(\text{opt})} = \frac{S_y S_y \lambda_{120} - S_{xy} \lambda_{100}}{S_x^2 (\lambda_{40} - 1) - \lambda_{100}}$, $\beta_{3(\text{opt})} = \frac{S_y}{S_x^2 (\lambda_{40} - 1)}$. |
| 5   | $\beta_1(\text{opt}) = \frac{S_{xy} (\lambda_{40} - 1) - S_x S_y \lambda_{100}}{S_x^2 (\lambda_{40} - 1) - \lambda_{100}}$, and $\beta_{2(\text{opt})} = \frac{S_y S_y \lambda_{120} - S_{xy} \lambda_{100}}{S_x^2 (\lambda_{40} - 1) - \lambda_{100}}$, $\beta_{3(\text{opt})} = \frac{S_y}{S_x^2 (\lambda_{40} - 1)}$. |
| 6   | $\beta_1(\text{opt}) = \frac{S_{xy} (\lambda_{40} - 1) - S_x S_y \lambda_{100}}{S_x^2 (\lambda_{40} - 1) - \lambda_{100}}$, and $\beta_{2(\text{opt})} = \frac{S_y S_y \lambda_{120} - S_{xy} \lambda_{100}}{S_x^2 (\lambda_{40} - 1) - \lambda_{100}}$, $\beta_{3(\text{opt})} = \frac{S_y}{S_x^2 (\lambda_{40} - 1)}$. |
| 7   | $\beta_1(\text{opt}) = \frac{S_{xy} (\lambda_{40} - 1) - S_x S_y \lambda_{100}}{S_x^2 (\lambda_{40} - 1) - \lambda_{100}}$, and $\beta_{2(\text{opt})} = \frac{S_y S_y \lambda_{120} - S_{xy} \lambda_{100}}{S_x^2 (\lambda_{40} - 1) - \lambda_{100}}$, $\beta_{3(\text{opt})} = \frac{S_y}{S_x^2 (\lambda_{40} - 1)}$. |
| 8   | $\beta_1(\text{opt}) = \frac{w_1}{t_1} = \beta_2(\text{opt}) = \frac{w_2}{t_2} = \beta_3(\text{opt}) = \frac{w_3}{t_3} = B_3$. |
Table 7: Mean square error equations of the proposed estimators.

| \( w \) | \( \text{MSE}(\hat{y}_{k, \text{est} \min}) \) |
|---|---|
| 1 | \( \theta_{n, N} S_y^2 + \frac{1}{r} \left[ Y^2 C^2_{\theta_1} \left( 1 + \frac{(N-1)C^2_{\theta_1}}{N} \right) + \frac{\theta^2_{\theta_1}}{\theta^2_{\theta_2}} C^2_{\theta_2} \right] \). |
| 2 | \( \theta_{n, N} S_y^2 + \frac{1}{r} \left[ Y^2 C^2_{\theta_1} \left( 1 + \frac{(N-1)C^2_{\theta_1}}{N} \right) + \frac{\theta^2_{\theta_1}}{\theta^2_{\theta_2}} C^2_{\theta_2} + \theta_{r, n} S_y^2 \left( 1 - \rho^2_{xy} \right) \right] \). |
| 3 | \( \theta_{n, N} S_y^2 + \frac{1}{r} \left[ Y^2 C^2_{\theta_1} \left( 1 + \frac{(N-1)C^2_{\theta_1}}{N} \right) + \frac{\theta^2_{\theta_1}}{\theta^2_{\theta_2}} C^2_{\theta_2} + \theta_{r, n} S_y^2 \left( 1 - \frac{\lambda^2_{120}}{(\lambda_{120} - 1)^2} \right) \right] \). |
| 4 | \( \theta_{n, N} S_y^2 + \frac{1}{r} \left[ Y^2 C^2_{\theta_1} \left( 1 + \frac{(N-1)C^2_{\theta_1}}{N} \right) + \frac{\theta^2_{\theta_1}}{\theta^2_{\theta_2}} C^2_{\theta_2} + \theta_{r, n} S_y^2 \left( 1 - \rho^2_{xy} - \frac{(\lambda_{120} - \rho_{x,y})(\lambda_{120})}{\lambda_{120} - 1} \right) \right] \). |
| 5 | \( \theta_{n, N} S_y^2 + \frac{1}{r} \left[ Y^2 C^2_{\theta_1} \left( 1 + \frac{(N-1)C^2_{\theta_1}}{N} \right) + \frac{\theta^2_{\theta_1}}{\theta^2_{\theta_2}} C^2_{\theta_2} + \theta_{r, n} S_y^2 \left( 1 - \frac{\rho^2_{x,y} - 2\rho_{x,y}\rho_{x,y}}{1 - \rho^2_{x,y}} \right) \right] \). |
| 6 | \( \theta_{n, N} S_y^2 + \frac{1}{r} \left[ Y^2 C^2_{\theta_1} \left( 1 + \frac{(N-1)C^2_{\theta_1}}{N} \right) + \frac{\theta^2_{\theta_1}}{\theta^2_{\theta_2}} C^2_{\theta_2} + \theta_{r, n} S_y^2 \left( 1 - \rho^2_{xy} - \frac{\lambda^2_{120}}{(\lambda_{120} - 1)^2} \right) \right] \). |
| 7 | \( \theta_{n, N} S_y^2 + \frac{1}{r} \left[ Y^2 C^2_{\theta_1} \left( 1 + \frac{(N-1)C^2_{\theta_1}}{N} \right) + \frac{\theta^2_{\theta_1}}{\theta^2_{\theta_2}} C^2_{\theta_2} + \theta_{r, n} S_y^2 \left( 1 - \rho^2_{xy} - \frac{\lambda^2_{120}}{(\lambda_{120} - 1)^2} \right) \right] \). |
| 8 | \( \theta_{n, N} S_y^2 + \frac{1}{r} \left[ Y^2 C^2_{\theta_1} \left( 1 + \frac{(N-1)C^2_{\theta_1}}{N} \right) + \frac{\theta^2_{\theta_1}}{\theta^2_{\theta_2}} C^2_{\theta_2} + \theta_{r, n} S_y^2 + B^2_{1, y, x} + B^2_{2, x, y} + B^2_{3, x, y} (\lambda_{120} - 1) + \frac{B^2_{1, y, x}}{2} - 2B_{1}S_{xy} - 2B_{2}S_{x, y} + 2B_{1}B_{2}S_{x, y}^{2} + 2B_{1}B_{3}S_{x, y} + 2B_{2}B_{3}S_{x, y}^{2} \right] \). |
Table 8: PRE(\(\cdot\)) of generalized ratio and regression estimator.

| \(r\) | \(C_{\theta_1}\) | \(C_{\theta_2}\) | \(\theta_2^2/\theta_1^2\) | \(RR\) | PRE(R) | PRE(Re) |
|-----|--------------------|--------------------|-------------------|--------|--------|--------|
| 10  | 0.1                | 0.1                | 0.5               | 6552.5720 | 1127.1130 | 1069.7540 |
|     |                    |                    | 1.0               | 6552.5730 | 1127.1110 | 1069.7530 |
|     |                    |                    | 1.5               | 6552.5740 | 1127.1090 | 1069.7510 |
|     | 0.2                |                    | 0.5               | 1069.7530 | 1127.1110 | 1069.7530 |
|     |                    |                    | 1.0               | 6552.5730 | 1127.1060 | 1069.7480 |
|     |                    |                    | 1.5               | 6552.5760 | 1127.0970 | 1069.7400 |
| 20  | 0.1                | 0.1                | 0.5               | 6552.5810 | 673.1071 | 653.5542 |
|     |                    |                    | 1.0               | 7013.1150 | 673.1067 | 653.5538 |
|     |                    |                    | 1.5               | 7013.1170 | 673.1060 | 653.5532 |
|     | 0.2                |                    | 0.5               | 7013.1150 | 673.1067 | 653.5538 |
|     |                    |                    | 1.0               | 7013.1180 | 673.1050 | 653.5522 |
|     |                    |                    | 1.5               | 7013.1230 | 673.1023 | 653.5497 |
| 30  | 0.1                | 0.1                | 0.5               | 2819.2100 | 461.3460 | 636.1122 |
|     |                    |                    | 1.0               | 2819.2110 | 461.3458 | 636.1118 |
|     |                    |                    | 1.5               | 2819.2110 | 461.3455 | 636.1110 |
|     | 0.2                |                    | 0.5               | 2819.2110 | 654.5258 | 636.1118 |
|     |                    |                    | 1.0               | 2819.2120 | 654.5239 | 636.1099 |
|     |                    |                    | 1.5               | 2819.2150 | 654.5207 | 636.1069 |
|     | 0.2                |                    | 0.5               | 3049.4820 | 461.3460 | 452.8047 |
|     |                    |                    | 1.0               | 3049.4820 | 461.3458 | 452.8045 |
|     |                    |                    | 1.5               | 3049.4830 | 461.3455 | 452.8041 |
|     | 0.2                |                    | 0.5               | 3049.4820 | 461.3458 | 452.8045 |
|     |                    |                    | 1.0               | 3049.4840 | 461.3450 | 452.8037 |
|     |                    |                    | 1.5               | 3049.4860 | 461.3437 | 452.8024 |
|     | 0.2                |                    | 0.5               | 1574.7570 | 413.8498 | 407.1786 |
|     |                    |                    | 1.0               | 1574.7570 | 413.8496 | 407.1784 |
|     |                    |                    | 1.5               | 1574.7570 | 413.8493 | 407.1781 |
|     | 0.2                |                    | 0.5               | 1574.7570 | 413.8496 | 407.1784 |
|     |                    |                    | 1.0               | 1574.7580 | 413.8488 | 407.1777 |
|     |                    |                    | 1.5               | 1574.7600 | 413.8474 | 407.1763 |
|     | 0.2                |                    | 0.5               | 1728.2710 | 323.6291 | 319.8946 |
|     |                    |                    | 1.0               | 1728.2710 | 323.6290 | 319.8945 |
|     |                    |                    | 1.5               | 1728.2710 | 323.6288 | 319.8944 |
|     | 0.2                |                    | 0.5               | 1728.2710 | 323.6290 | 319.8945 |
|      |    | 0.1 | 0.1 | 0.5 | 1.0 | 1.5 | 0.2 | 0.5 | 1.0 | 1.5 | 0.2 | 0.1 | 0.5 | 1.0 | 1.5 |
|-----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 40  |    |     |     |     | 1.0 | 1728.2720 | 323.6286 | 319.8941 | 1.5 | 1728.2740 | 323.6279 | 319.8934 | 1.5 | 952.5302 | 268.0097 | 265.6795 |
|     |    |     |     |     | 0.5 | 952.5297 | 268.0099 | 265.6797 | 1.0 | 265.6796 | 268.0098 | 952.5299 | 1.5 | 952.5302 | 268.0097 | 265.6795 |
|     |    |     |     |     | 0.2 | 952.5306 | 268.0095 | 265.6793 | 1.0 | 952.5319 | 268.0089 | 265.6787 | 1.5 | 952.5302 | 268.0097 | 265.6795 |
| 0.2 | 0.1| 0.1 | 0.5 | 1067.6650 | 226.9002 | 225.4070 | 1.0 | 1067.6660 | 226.9001 | 225.4069 | 1.5 | 1067.6660 | 226.9001 | 225.4068 |
|     |    |     |     | 0.5 | 1067.6660 | 226.9001 | 225.4069 | 1.0 | 1067.6660 | 226.9001 | 225.4068 | 1.5 | 1067.6660 | 226.9001 | 225.4068 |
|     |    |     |     | 0.2 | 1067.6660 | 226.9001 | 225.4069 | 1.0 | 1067.6660 | 226.8999 | 225.4067 | 1.5 | 1067.6680 | 226.8996 | 225.4064 |
| 50  |    |     |     |     | 0.5 | 579.1935 | 170.1783 | 169.5572 | 1.0 | 579.1937 | 170.1783 | 169.5571 | 1.5 | 579.1937 | 170.1783 | 169.5571 |
|     |    |     |     |     | 0.2 | 579.1937 | 170.1783 | 169.5571 | 1.0 | 579.1943 | 170.1782 | 169.5570 | 1.5 | 579.1953 | 170.1780 | 169.5568 |
|     |    |     |     |     | 0.2 | 579.1937 | 170.1783 | 169.5571 | 1.0 | 579.1925 | 155.2310 | 154.7847 | 1.5 | 579.1932 | 155.2308 | 154.7845 |
Table 9: Descriptive statistics of the results.

| $r$ | Freq. | Mean | Std. | Min. | Med. | Max. |
|-----|-------|------|------|------|------|------|
| PRE(R) |     |      |      |      |      |      |
| 10  | 12   | 900.1067 | 237.0949 | 673.1023 | 900.1021 | 1127.1130 |
| 20  | 12   | 509.6400 | 87.3680  | 461.3437 | 461.3458 | 654.5258 |
| 30  | 12   | 368.7389 | 47.1160  | 323.6279 | 368.7383 | 413.8498 |
| 40  | 12   | 247.4548 | 21.4688  | 226.8996 | 247.4546 | 268.0099 |
| 50  | 12   | 162.7046 | 7.8060   | 155.2308 | 162.7045 | 170.1783 |
| PRE(Re) |     |      |      |      |      |      |
| 10  | 12   | 861.6513 | 217.3518 | 653.5497 | 861.6471 | 1069.7540 |
| 20  | 12   | 544.4573 | 95.7288  | 452.8024 | 544.4558 | 636.1122 |
| 30  | 12   | 363.5361 | 45.5824  | 319.8934 | 363.5355 | 407.1786 |
| 40  | 12   | 302.7806 | 205.5982 | 225.4064 | 245.5429 | 952.5299 |
| 50  | 12   | 162.1709 | 7.7146   | 154.7845 | 162.1708 | 169.5572 |

References

Ahmed, M. S., Al-Titi, O., Al-Rawi, Z., and Abu-Dayyeh, W. (2006), Estimation of a population mean using different imputation methods. *Statistics in Transition* 7(6), 1247, 1264.

Bar-Lev, S. K., Bobovitch, E., and Boukai, B. (2004), A note on randomized response models for quantitative data. *Metrika*, 60(3), 255-260.

Chaudhuri, A., and Adhikary, A. K. (1990), Variance estimation with randomized response. *Communications in Statistics-Theory and Methods*, 19 (3), 1119-1125.

Diana, G., and Perri, P. F. (2010), New scrambled response models for estimating the mean of a sensitive quantitative character. *Journal of Applied Statistics*, 37(11), 1875-1890.

Eichhorn, B. H., and Hayre, L. S. (1983), Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and inference*, 7(4), 307-316.

Folsom, R. E., Greenberg, B. G., Horvitz, D. G., and Abernathy, J. R. (1973), The two
alternate questions randomized response model for human surveys. *Journal of the American Statistical Association, 68*(343), 525-530.

Gjestvang, C. R., and Singh, S. (2006), A new randomized response model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68*(3), 523-530.

Gjestvang, C. R., and Singh, S. (2009), An improved randomized response model: Estimation of mean. *Journal of Applied Statistics, 36*(12), 1361-1367.

Greenberg, B. G., Abul-Ela, A. L. A., Simmons, W. R., and Horvitz, D. G. (1969), The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association, 64*(326), 520-539.

Greenberg, B. G., Kuebler Jr, R. R., Abernathy, J. R., and Horvitz, D. G. (1971), Application of the randomized response technique in obtaining quantitative data. *Journal of the American Statistical Association, 66*(334), 243-250.

Gupta, S., Mehta, S., Shabbir, J., and Dass, B. K. (2013), Generalized scrambling in quantitative optional randomized response models. *Communications in Statistics-Theory and Methods, 42*(22), 4034-4042.

Gupta, S., Shabbir, J., and Sehra, S. (2010), Mean and sensitivity estimation in optional randomized response models. *Journal of Statistical Planning and Inference, 140*(10), 2870-2874.

Haq, A., Khan, M., and Hussain, Z. (2017), A new estimator of finite population mean based on the dual use of the auxiliary information. *Communications in Statistics-Theory and Methods, 46*(9), 4425-4436.

Heitjan, D. F., and Basu, S. (1996), Distinguishing missing at random and missing completely at random. *The American Statistician, 50*(3), 207-213.

Mangat, N. S., and Singh, R. (1990), An alternative randomized response procedure. *Biometrika, 77*(2), 439-442.

Mohamed, C., Sedory, S. A., and Singh, S. (2017), Imputation using higher order moments of an auxiliary variable. *Communications in Statistics-Simulation and Computation, 46*(8), 6588-6617.

Moors, J. J. A. (1971), Optimization of the unrelated question randomized response model. *Journal of the American Statistical Association, 66*(335), 627-629.
Murthy, M. N. (1967), Sampling theory and methods. *Sampling theory and methods*.

Rosenfeld, B., Imai, K., and Shapiro, J. N. (2016), An empirical validation study of popular survey methodologies for sensitive questions. *American Journal of Political Science, 60*(3), 783-802.

Rubin, D. B. (1976), Inference and missing data. *Biometrika, 63*(3), 581-592.

Singh, G. N., Priyanka, K., Kim, J. M., and Singh, S. (2010), Estimation of population mean using imputation techniques in sample surveys. *Journal of the Korean Statistical Society, 39*(1), 67-74.

Singh, S., and Deo, B. (2003), Imputation by power transformation. *Statistical Papers, 44*(4), 555-579.

Warner, S. L. (1965), Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association, 60*(309), 63-69.
Appendix

For evaluating the bias and variance of the estimators, we define following useful

\[ E = \text{To the first order approximation, we have} \]

\[ e_0 = \frac{r}{R} - 1, e_1 = \frac{r}{X} - 1, e_2 = \frac{r}{X} - 1, e_3 = \frac{r}{X} - 1, e_4 = \frac{r}{X} - 1, e_5 = \frac{r}{X} - 1, \]

\[ e_6 = \frac{r}{R} - 1, e_7 = \frac{r}{R} - 1, E(e_i) = 0, (i = 0, 1, 2, 3, 4, 5, 6, 7), \]

To the first order approximation, we have

\[ E\left(\frac{e_0^2}{n}\right) = \theta_{r,N}C_y^2 + \frac{1}{N} \left[ C_0^2 \left\{ 1 + \frac{(N-1)C_y^2}{N} \right\} + \frac{\theta_0^2}{\bar{y}^2} \right], \]

\[ E\left(\frac{e_1^2}{2}\right) = \theta_{r,N}C_x^2, E\left(\frac{e_2^2}{2}\right) = \theta_{n,N}C_x^2, E\left(\frac{e_3^2}{3}\right) = \theta_{r,N}(\lambda_{040} - 1), E\left(\frac{e_4^2}{4}\right) = \theta_{n,N}(\lambda_{040} - 1), \]

\[ E\left(\frac{e_5^2}{6}\right) = \theta_{r,N}C_x^2, E\left(\frac{e_6^2}{2}\right) = \theta_{n,N}(\lambda_{003} - 1), E\left(\frac{e_7^2}{7}\right) = \theta_{n,N}(\lambda_{003} - 1), \]

where

\[ \tilde{\tau} = \frac{1}{N} \sum_{j=1}^{N} \tau_j, C_\tilde{\tau} = \frac{\bar{y}^2}{\bar{y}^2}, \rho_{\tilde{\tau}\psi} = \frac{S_{\psi\tilde{\tau}}}{\bar{y}^2}, C_\tilde{\psi} = \left( \frac{\bar{\psi}}{\bar{\psi}} \right)^2, C_{\tilde{\mu}} = \left( \frac{\bar{\mu}}{\bar{\mu}} \right)^2, R = \frac{R}{\bar{R}}, \]

\[ R' = \frac{\bar{y}}{R} S_{\tilde{\tau}\psi} = \frac{1}{N} \sum_{j=1}^{N} (1 - \tilde{\tau})(\tilde{\psi}_j - \bar{\psi}), \theta_{r,N} = \left( \frac{1}{1 - \tilde{\tau}} \right), \theta_{n,N} = \left( \frac{1}{1 - \tilde{\tau}} \right), \]

\[ \theta_{r,N} = \left( \frac{1}{1 - \tilde{\tau}} \right), \mu_{abc} = \frac{1}{N-1} \sum_{j=1}^{N} (y_j - \bar{Y})(x_j - \bar{X})(r_j - R), \lambda_{abc} = \frac{\mu_{abc}}{\mu_{x\tilde{\tau}}^2 \mu_{y\tilde{\psi}}^2 \mu_{r\tilde{\mu}}^2}, \]

where \( \tilde{\tau} = R, X, Y \) and \( \psi = R, X, Y \).
## Table 10: P.R.E(i) of the special cases of generalized ratio method of imputation.

| $r$ | $C_{01}$ | $C_{02}$ | $\theta^2/\theta^2_{i-1}$ | $RR$ | PRE(2) | PRE(3) | PRE(4) | PRE(5) | PRE(6) | PRE(7) | PRE(8) |
|-----|----------|----------|---------------------------|------|--------|--------|--------|--------|--------|--------|--------|
| 10  | 0.1      | 0.5      |                           | 6552.5720 | 784.7886 | 132.2070 | 290.4700 | 67.8920 | 343.3627 | 993.4518 | 2099.6610 |
|     | 0.2      | 0.5      |                           | 1069.7530 | 784.7880 | 132.2070 | 290.4700 | 67.8920 | 343.3626 | 993.4508 | 2099.6600 |
|     | 0.2      | 0.5      |                           | 413.1450  | 589.7870 | 132.2070 | 290.4698 | 67.8920 | 343.3625 | 993.4468 | 2099.6370 |
|     | 20       | 0.1      |                           | 2819.2100 | 529.7668 | 129.4688 | 258.1774 | 69.3543 | 293.2027 | 610.2027 | 872.7863 |
|     | 0.2      | 0.5      |                           | 1574.7570 | 365.1765 | 125.4285 | 220.1187 | 71.7581 | 243.8033 | 397.5078 | 482.0815 |
|     | 0.2      | 0.5      |                           | 1728.2710 | 295.5593 | 122.6580 | 198.8972 | 73.6045 | 216.1887 | 314.4188 | 359.9275 |
|     | 30       | 0.1      |                           | 1574.7570 | 365.1765 | 125.4285 | 220.1187 | 71.7581 | 243.8033 | 397.5078 | 482.0815 |
|     | 0.2      | 0.5      |                           | 1574.7570 | 365.1765 | 125.4285 | 220.1187 | 71.7581 | 243.8033 | 397.5078 | 482.0812 |
|     | 0.2      | 0.5      |                           | 1574.7580 | 365.1757 | 125.4285 | 220.1185 | 71.7581 | 243.8030 | 397.5068 | 482.0801 |
|     | 0.2      | 0.5      |                           | 1574.7600 | 365.1747 | 125.4284 | 220.1182 | 71.7581 | 243.8026 | 397.5056 | 482.0781 |

---

Table continues with similar entries for different values of $r$. The table provides a comprehensive list of P.R.E(i) values for various scenarios with different parameters. The entries are formatted to show the effect of varying parameters on the imputation method's performance.
|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| 0.2 | 0.5 | 1728.2710 | 295.5592 | 122.6580 | 198.8972 | 73.6046 | 216.1887 | 314.4187 | 359.9274 |
| 1.0 | 1728.2720 | 295.5589 | 122.6580 | 198.8971 | 73.6046 | 216.1885 | 314.4183 | 359.9268 |
| 1.5 | 1728.2740 | 295.5583 | 122.6580 | 198.8969 | 73.6046 | 216.1883 | 314.4176 | 359.9259 |
| 0.2 | 0.1 | 1728.2720 | 295.5589 | 122.6580 | 198.8971 | 73.6046 | 216.1883 | 314.4176 | 359.9259 |
| 1.0 | 265.6796 | 250.1606 | 120.1321 | 182.1776 | 75.4528 | 195.1474 | 262.2386 | 289.9808 |
| 1.5 | 952.5302 | 250.1604 | 120.1320 | 182.1775 | 75.4528 | 195.1473 | 262.2384 | 289.9805 |
| 0.2 | 1.0 | 1728.2720 | 295.5589 | 122.6580 | 198.8971 | 73.6046 | 216.1883 | 314.4176 | 359.9259 |
| 1.0 | 265.6796 | 250.1606 | 120.1321 | 182.1776 | 75.4528 | 195.1474 | 262.2386 | 289.9808 |
| 1.5 | 952.5302 | 250.1604 | 120.1320 | 182.1775 | 75.4528 | 195.1473 | 262.2384 | 289.9805 |
| 0.2 | 1.5 | 1728.2740 | 295.5583 | 122.6580 | 198.8969 | 73.6046 | 216.1883 | 314.4176 | 359.9259 |
| 1.0 | 265.6796 | 250.1606 | 120.1321 | 182.1776 | 75.4528 | 195.1474 | 262.2386 | 289.9808 |
| 1.5 | 952.5302 | 250.1604 | 120.1320 | 182.1775 | 75.4528 | 195.1473 | 262.2384 | 289.9805 |

Table 11: P.R.E(i) of the special cases of generalized regression method of imputation.
| 0.2 | 0.1 | 0.5 | 6552.5810 | 662.1809 | 228.9495 | 368.6932 | 1187.1660 | 1139.5230 | 114.4118 | 1127.1060 |
| 1.0 | 7013.1100 | 662.1805 | 228.9495 | 368.6930 | 1187.1550 | 1139.5130 | 114.4117 | 1127.0970 |
| 1.5 | 7013.1170 | 662.1798 | 228.9494 | 368.6924 | 1187.1560 | 1139.5130 | 114.4117 | 1127.0970 |
| 0.2 | 0.5 | 6552.5810 | 662.1809 | 228.9495 | 368.6932 | 1187.1660 | 1139.5230 | 114.4118 | 1127.1060 |
| 1.0 | 7013.1100 | 662.1805 | 228.9495 | 368.6930 | 1187.1550 | 1139.5130 | 114.4117 | 1127.0970 |
| 1.5 | 7013.1170 | 662.1798 | 228.9494 | 368.6924 | 1187.1560 | 1139.5130 | 114.4117 | 1127.0970 |
| 0.2 | 0.1 | 0.5 | 2819.2100 | 644.2407 | 227.4964 | 310.1177 | 672.9058 | 658.3995 | 113.2643 | 654.5263 |
| 1.0 | 2819.2110 | 644.2403 | 227.4964 | 310.1176 | 672.9053 | 658.3990 | 113.2643 | 654.5258 |
| 1.5 | 2819.2110 | 644.2395 | 227.4963 | 310.1175 | 672.9045 | 658.3982 | 113.2643 | 654.5250 |
| 0.2 | 0.5 | 2819.2100 | 644.2407 | 227.4964 | 310.1177 | 672.9058 | 658.3995 | 113.2643 | 654.5263 |
| 1.0 | 2819.2120 | 644.2384 | 227.4962 | 310.1173 | 672.9033 | 658.3971 | 113.2643 | 654.5239 |
| 1.5 | 2819.2150 | 644.2353 | 227.4960 | 310.1167 | 672.8999 | 658.3938 | 113.2643 | 654.5207 |
| 0.2 | 0.1 | 0.5 | 3049.4820 | 456.5958 | 207.5177 | 267.6513 | 469.7061 | 463.1214 | 112.1411 | 461.3460 |
| 1.0 | 3049.4820 | 456.5956 | 207.5177 | 267.6512 | 469.7059 | 463.1212 | 112.1411 | 461.3458 |
| 1.5 | 3049.4830 | 456.5953 | 207.5176 | 267.6511 | 469.7055 | 463.1208 | 112.1411 | 461.3455 |
| 0.2 | 0.5 | 3049.4820 | 456.5958 | 207.5177 | 267.6512 | 469.7059 | 463.1212 | 112.1411 | 461.3458 |
| 1.0 | 3049.4840 | 456.5948 | 207.5176 | 267.6510 | 469.7050 | 463.1203 | 112.1411 | 461.3450 |
| 1.5 | 3049.4860 | 456.5935 | 207.5174 | 267.6506 | 469.7036 | 463.1190 | 112.1411 | 461.3437 |
| 0.2 | 0.1 | 0.5 | 952.5297 | 266.7197 | 184.1950 | 223.5201 | 327.2336 | 324.3975 | 110.5603 | 323.6279 |
| 1.0 | 952.5306 | 266.7196 | 184.1950 | 223.5200 | 327.2325 | 324.3974 | 110.5603 | 323.6273 |
| 1.5 | 952.5306 | 266.7195 | 184.1950 | 223.5200 | 327.2323 | 324.3972 | 110.5603 | 323.6270 |
| 0.2 | 0.5 | 952.5297 | 266.7197 | 184.1950 | 223.5201 | 327.2336 | 324.3975 | 110.5603 | 323.6279 |
| 1.0 | 952.5306 | 266.7196 | 184.1950 | 223.5200 | 327.2325 | 324.3974 | 110.5603 | 323.6273 |
| 1.5 | 952.5306 | 266.7195 | 184.1950 | 223.5200 | 327.2323 | 324.3972 | 110.5603 | 323.6270 |
|   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|
| 0.2 | 0.1 | 1.5 | 952.5319 | 266.7187 | 170.8462 | 200.5336 | 270.2453 | 268.4874 | 109.4873 | 268.0089 |
| 0.5 | 0.5 | 1067.6650 | 226.0742 | 158.7203 | 180.9197 | 228.3274 | 227.2061 | 108.3785 | 226.9002 |
| 1.0 | 1.0 | 1067.6660 | 226.0742 | 158.7203 | 180.9197 | 228.3273 | 227.2060 | 108.3785 | 226.9001 |
| 1.5 | 1.5 | 1067.6660 | 226.0741 | 158.7202 | 180.9196 | 228.3272 | 227.2059 | 108.3785 | 226.9001 |
| 0.2 | 0.2 | 0.1 | 0.5 | 579.1935 | 169.8352 | 137.5117 | 149.2074 | 170.7686 | 170.3051 | 106.0448 | 170.1783 |
| 1.0 | 1.0 | 579.1937 | 169.8352 | 137.5117 | 149.2074 | 170.7686 | 170.3051 | 106.0448 | 170.1783 |
| 1.5 | 1.5 | 579.1937 | 169.8351 | 137.5117 | 149.2073 | 170.7686 | 170.3051 | 106.0448 | 170.1783 |
| 0.2 | 0.2 | 0.1 | 0.5 | 671.3021 | 154.9846 | 130.7805 | 139.7705 | 155.6545 | 155.3220 | 105.1725 | 155.2310 |
| 1.0 | 1.0 | 671.3022 | 154.9846 | 130.7805 | 139.7705 | 155.6545 | 155.3220 | 105.1725 | 155.2310 |
| 1.5 | 1.5 | 671.3025 | 154.9845 | 130.7805 | 139.7705 | 155.6544 | 155.3220 | 105.1725 | 155.2310 |
| 0.2 | 0.2 | 0.1 | 0.5 | 671.3022 | 154.9846 | 130.7805 | 139.7705 | 155.6545 | 155.3220 | 105.1725 | 155.2310 |
| 1.0 | 1.0 | 671.3028 | 154.9845 | 130.7805 | 139.7704 | 155.6544 | 155.3219 | 105.1725 | 155.2309 |
| 1.5 | 1.5 | 671.3058 | 154.9843 | 130.7804 | 139.7704 | 155.6542 | 155.3218 | 105.1725 | 155.2308 |
Figure 1: Illustration of data imputation scheme.