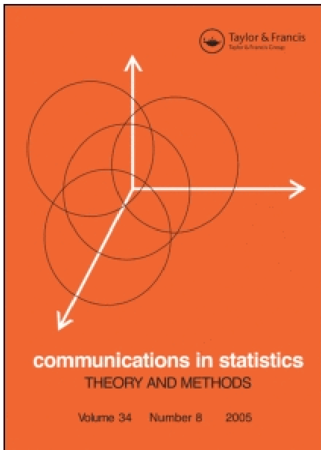

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597238

Estimators for the Population Mean in the Case of Missing Data

Cem Kadilar [a]; Hulya Cingi [a]
[a] Department of Statistics, Hacettepe University, Ankara, Turkey

Online Publication Date: 01 January 2008

To cite this Article: Kadilar, Cem and Cingi, Hulya (2008) 'Estimators for the Population Mean in the Case of Missing Data', Communications in Statistics - Theory and Methods, 37:14, 2226 — 2236

To link to this article: DOI: 10.1080/03610920701855020
URL: http://dx.doi.org/10.1080/03610920701855020

Communications in Statistics—Theory and Methods, 37: 2226–2236, 2008

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920701855020

Sampling Theory

Estimators for the Population Mean in the Case of Missing Data

CEM KADILAR AND HULYA CINGI

Department of Statistics, Hacettepe University, Ankara, Turkey

We propose a class of estimators for the population mean when there are missing data in the data set. Obtaining the mean square error equations of the proposed estimators, we show the conditions where the proposed estimators are more efficient than the sample mean, ratio-type estimators, and the estimators in Singh and Horn (2000) *and Singh and Deo* (2003) *in the case of missing data. These conditions are also supported by a numerical example.*

Keywords Efficiency; Missing data; Ratio-type estimator; Regression estimator; Simple random sampling.

Mathematics Subject Classification 62D05; 62G05.

1. Introduction

Missing data is a common problem in sample surveys and statisticians have recognized that statistical inference can be spoiled in the presence of non response. Therefore, there are some recent studies on this important topic such as Singh and Horn (2000), Singh and Deo (2003), Rueda and Gonzalez (2004), Rueda et al. (2005), etc. In all of these articles, the purpose is to estimate the population mean efficiently when some observations are missing from the sample.

It is well known that under the mean method of imputation, the sample mean and its variance can be given by:

$$\bar{y}_s = \frac{1}{r}\sum_{i=1}^{r} y_i, \tag{1}$$

$$V(\bar{y}_s) = \left(\frac{1}{r} - \frac{1}{N}\right)S_y^2, \tag{2}$$

Received May 3, 2007; Accepted December 10, 2007
Address correspondence to Cem Kadilar, Department of Statistics, Hacettepe University, Beytepe, Ankara 06800, Turkey; E-mail: kadilar@hacettepe.edu.tr

respectively. Here, r is the number of responding units out of sampled n units drawn from the finite population of size, N, by simple random sampling without replacement (SRSWOR) and S_y^2 is the population variance of the study variable. Note that the sample mean is an unbiased estimator of the population mean so $V(\bar{y}_s) = MSE(\bar{y}_s)$.

It is also known that under the ratio method of imputation, the ratio estimator for the population mean, its bias, and mean square error (MSE) in the case of missing data can be given by:

$$\bar{y}_r = \frac{\bar{y}_s}{\bar{x}_s}\bar{x}, \tag{3}$$

$$\begin{aligned} B(\bar{y}_r) &\cong \left(\frac{1}{r} - \frac{1}{n}\right)\overline{Y}\left(C_x^2 - \rho C_y C_x\right), \\ MSE(\bar{y}_r) &\cong \left(\frac{1}{r} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{r} - \frac{1}{n}\right)\left(R^2 S_x^2 - 2RS_{xy}\right), \end{aligned} \tag{4}$$

respectively. Note that we implicitly assume missing completely at random (MCAR) in the present investigation as in Singh and Horn (2000). Here $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ is the sample mean of the auxiliary variable, $\bar{x}_s = \frac{\sum_{i=1}^{r} x_i}{r}$ is the sample mean of the auxiliary variable under the mean method of imputation, $C_x = \frac{S_x}{\overline{X}}$ and $C_y = \frac{S_y}{\overline{Y}}$ are the population coefficients of variation of auxiliary and study variables, respectively, ρ is the coefficient of correlation between the auxiliary and the study variables, $R = \frac{\overline{Y}}{\overline{X}}$ is the population ratio, $\overline{X}$ and $\overline{Y}$ are the population means of the auxiliary and the study variables, respectively, S_x^2 is the population variance of the auxiliary variable, S_{xy} is the population covariance between the auxiliary and the study variables.

Comparing (2) and (4), one can easily see that the ratio method of imputation is more efficient than the mean method of imputation when:

$$\begin{aligned} R &< 2B, \quad \text{for } R > 0 \\ R &> 2B, \quad \text{for } R < 0 \end{aligned} \tag{5}$$

where $B = \frac{S_{xy}}{S_x^2}$.

In order to have efficient population mean estimator, Singh and Horn (2000) and Singh and Deo (2003) proposed the following estimators, respectively:

$$\bar{y}_{SH} = \alpha\,\bar{y}_s + (1-\alpha)\bar{y}_s\frac{\bar{x}}{\bar{x}_s}, \tag{6}$$

$$\bar{y}_{SD} = \bar{y}_s\left(\frac{\bar{x}}{\bar{x}_s}\right)^{\gamma}, \tag{7}$$

where α is a suitable constant. The optimal values α and γ, which make the MSE minimum, are computed by:

$$\alpha^* = 1 - \rho\frac{C_y}{C_x} \quad \text{and} \quad \gamma^* = \rho\frac{C_y}{C_x},$$

respectively. Using these optimal values, the minimum MSE equations can be obtained. Singh and Horn (2000) and Singh and Deo (2003) obtained the biases and minimum MSE equations of these estimators as follows:

$$B(\bar{y}_{SH}) \cong \left(1-\alpha\right)\left(\frac{1}{r}-\frac{1}{n}\right)\overline{Y}(C_x^2-\rho C_y C_x),$$

$$B(\bar{y}_{SD}) \cong \left(\frac{1}{r}-\frac{1}{n}\right)\overline{Y}\left(\frac{\alpha(\alpha-1)}{2}C_x^2-\alpha\rho C_y C_x\right)$$

and

$$MSE_{\min}(\bar{y}_{SH}) \cong MSE(\bar{y}_r)-\left(\frac{1}{r}-\frac{1}{n}\right)\left(1-\rho\frac{C_y}{C_x}\right)^2\overline{Y}^2C_x^2, \tag{8}$$

$$MSE_{\min}(\bar{y}_{SD}) \cong MSE(\bar{y}_r)-\left(\frac{1}{r}-\frac{1}{n}\right)S_x^2(B-R)^2, \tag{9}$$

respectively. It is clearly seen that these two estimators are always more efficient than the ratio estimator, $\bar{y}_r$. We found that the MSE equations, given in (8) and (9), are equal to each other (for details see Appendix).

2. Suggested Estimators

Kadilar and Cingi (2004) suggested the following estimator for the population mean in the simple random sampling:

$$\bar{y}_{KC}=\frac{\bar{y}+b(\overline{X}-\bar{x})}{\bar{x}}\overline{X}, \tag{10}$$

where $b=\frac{s_{xy}}{s_x^2}$ is the regression coefficient computed by the least square method (for details see Kadilar and Cingi, 2005). Here, s_x^2 is the sample variance of the auxiliary variable and s_{xy} is the sample covariance between the auxiliary and the study variables. Note that Ray and Singh (1981) examined the power transformation of the estimator in (10).

When we modify the estimator in (10) to account for the ratio imputation method in (3) and the mean method of imputation in (1), we propose first estimator as follows:

$$\bar{y}_{pr1}=\frac{\bar{y}_s+b(\overline{X}-\bar{x})}{\bar{x}}\overline{X}, \tag{11}$$

where $\overline{X}=\frac{\sum_{i=1}^{N}x_i}{N}$.

Theorem 2.1. *The bias of the proposed estimator in* (11) *is given by*:

$$B(\bar{y}_{pr1}) \cong \left(\frac{1}{n}-\frac{1}{N}\right)\overline{Y}C_x^2. \tag{12}$$

Proof. The bias of the proposed estimator in (11) can be obtained using the second-degree approximation, $0(n^{-2})$, in the Taylor series method defined by:

$$B(\bar{y}_{pr}) \cong \frac{1}{2}\left[\sum_{i=1}^{p}\sum_{j=1}^{p}d_{ij}E(\hat{\eta}_i-\eta_i)(\hat{\eta}_j-\eta_j)\right], \tag{13}$$

where p is the number of parameters, $p = 2$ for this proposed estimator, $d_{ij} = \frac{\partial^2 h(t)}{\partial \hat{\eta}_i \partial \hat{\eta}_j}\Big|_{t=T}$, $h(t) = h(\bar{y}_s, \bar{x}) = \bar{y}_{pr1}$, $\hat{\eta}_1 = \bar{y}_s$, $\hat{\eta}_2 = \bar{x}$, and $\eta_1 = \overline{Y}$, $\eta_2 = \overline{X}$ (Wolter, 1985). By these definitions, we can re-write (13) as:

$$B(\bar{y}_{pr1}) \cong \frac{1}{2}\{d_{11}V(\bar{y}_s) + d_{12}cov(\bar{x}, \bar{y}_s) + d_{21}cov(\bar{y}_s, \bar{x}) + d_{22}V(\bar{x})\} \tag{14}$$

where

$$V(\bar{y}_s) = \left(\frac{1}{r} - \frac{1}{N}\right)S_y^2, \tag{15}$$

$$V(\bar{x}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_x^2, \tag{16}$$

$$cov(\bar{x}, \bar{y}_s) = cov(\bar{y}_s, \bar{x}) = \left(\frac{1}{n} - \frac{1}{N}\right)\rho S_y S_x, \tag{17}$$

(Singh and Deo, 2003) and

$$d_{11} = 0, \quad d_{12} = d_{21} = -\frac{1}{\overline{X}}, \quad \text{and} \quad d_{22} = \frac{2}{\overline{X}}(R + B) \tag{18}$$

Substituting (15)–(18) in (10), we get (12). Hence, the theorem. □

We would like to note that $E(b) = B$ (see Cochran, 1977).

Theorem 2.2. *The MSE of the proposed estimator in* (11) *is given by*:

$$MSE(\bar{y}_{pr1}) \cong \left(\frac{1}{r} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{n} - \frac{1}{N}\right)S_x^2(R^2 - B^2). \tag{19}$$

Proof. Using the first degree approximation, $0(n^{-1})$, in the Taylor series method, we can write:

$$MSE(\bar{y}_{pr1}) \cong \frac{\overline{Y}^2 + 2B\overline{XY} + B^2\overline{X}^2}{\overline{X}^2}V(\bar{x}) - \frac{2\bar{Y} + 2B\overline{X}}{\overline{X}}cov(\bar{x}, \bar{y}_s) + V(\bar{y}_s) \tag{20}$$

(see Kadilar and Cingi, 2004). Substituting (15)–(17) in (20), we get (19). Hence, the theorem. □

We would like to remark that $\rho S_y = BS_x$ (Cingi, 1994).

We propose second estimator as follows:

$$\bar{y}_{pr2} = \frac{\bar{y}_s + b(\overline{X} - \bar{x}_s)}{\bar{x}_s}\overline{X}. \tag{21}$$

Theorem 2.3. *The bias of the proposed estimator in* (21) *is given by*:

$$B(\bar{y}_{pr2}) \cong \left(\frac{1}{r} - \frac{1}{N}\right)\overline{Y}C_x^2. \tag{22}$$

Proof. The only difference between first and second proposed estimators is to replace $\bar{x}$ with $\bar{x}_s$ so we can write (14) as:

$$B(\bar{y}_{pr2}) \cong \frac{1}{2}\{d_{11}V(\bar{y}_s) + d_{12}cov(\bar{x}_s, \bar{y}_s) + d_{21}cov(\bar{y}_s, \bar{x}_s) + d_{22}V(\bar{x}_s)\} \tag{23}$$

where

$$V(\bar{x}_s) = \left(\frac{1}{r} - \frac{1}{N}\right)S_x^2 \tag{24}$$

$$cov(\bar{x}_s, \bar{y}_s) = cov(\bar{y}_s, \bar{x}_s) = \left(\frac{1}{r} - \frac{1}{N}\right)\rho S_y S_x. \tag{25}$$

Substituting (15), (18), (24), and (25) in (23), we get (22). Hence, the theorem. □

It is worth pointing out that the only difference between the biases of first and second proposed estimators is to replace n with r.

Theorem 2.4. *The MSE of the proposed estimator in* (21) *is given by*:

$$MSE(\bar{y}_{pr2}) \cong \left(\frac{1}{r} - \frac{1}{N}\right)(S_y^2 - BS_{xy} + R^2S_x^2). \tag{26}$$

Proof. For second proposed estimator, we can write (20) as:

$$MSE(\bar{y}_{pr2}) \cong \frac{\overline{Y}^2 + 2B\overline{XY} + B^2\overline{X}^2}{\overline{X}^2}V(\bar{x}_s) - \frac{2\overline{Y} + 2B\overline{X}}{\overline{X}}cov(\bar{x}_s, \bar{y}_s) + V(\bar{y}_s). \tag{27}$$

Substituting (15), (24), and (25) in (27), we get (26). Note that $BS_{xy} = B^2S_x^2$ (Cingi, 1994). □

We would like to remark that when n is replaced with r in (19), we get (26).

If the population mean of the auxiliary variable, $\overline{X}$, is unknown, we propose third estimator as:

$$\bar{y}_{pr3} = \frac{\bar{y}_s + b(\bar{x} - \bar{x}_s)}{\bar{x}_s}\bar{x}. \tag{28}$$

Theorem 2.5. *The bias of the proposed estimator in* (28) *is given by*:

$$B(\bar{y}_{pr3}) \cong \left(\frac{1}{r} - \frac{1}{n}\right)\overline{Y}\rho C_y C_x. \tag{29}$$

Proof. When $p = 3$ in (13), we can write:

$$\begin{aligned} B(\bar{y}_{pr3}) \cong \frac{1}{2}\{&d_{11}V(\bar{y}_s) + d_{21}cov(\bar{x}, \bar{y}_s) + d_{31}cov(\bar{x}_s, \bar{y}_s) + d_{12}cov(\bar{y}_s, \bar{x}) \\ &+ d_{22}V(\bar{x}) + d_{32}cov(\bar{x}_s, \bar{x}) + d_{13}cov(\bar{y}_s, \bar{x}_s) \\ &+ d_{23}cov(\bar{x}, \bar{x}_s) + d_{33}V(\bar{x}_s)\} \end{aligned} \tag{30}$$

where

$$cov(\bar{x}_s, \bar{x}) = cov(\bar{x}, \bar{x}_s) = \left(\frac{1}{n} - \frac{1}{N}\right)S_x^2, \tag{31}$$

(Singh and Deo, 2003)

$$d_{11} = 0, \quad d_{22} = \frac{2}{\overline{X}}(R + B), \quad d_{33} = 0, \quad d_{12} = d_{21} = -\frac{1}{\overline{X}},$$
$$d_{31} = d_{13} = \frac{1}{\overline{X}}, \quad d_{32} = d_{23} = -\frac{\overline{Y}}{\overline{X}^2} - \frac{B}{\overline{X}}. \tag{32}$$

Substituting (15)–(17), (24)–(25), and (31)–(32) in (30), we get (29). □

Theorem 2.6. *The MSE of the proposed estimator in* (28) *is given by*:

$$MSE(\bar{y}_{pr3}) \cong \left(\frac{1}{r} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{r} - \frac{1}{n}\right)\left[(R + B)^2 S_x^2 - 2(R + B)S_{xy}\right]. \tag{33}$$

Proof. The MSE of the proposed estimator can be found using the first-degree approximation in the Taylor series method defined by:

$$MSE(\bar{y}_{pr3}) \cong \boldsymbol{d}\Sigma\boldsymbol{d}', \tag{34}$$

where

$$\boldsymbol{d} = \left[\frac{\partial h(a, b, c)}{\partial a}\Big|_{\overline{Y},\overline{X}} \quad \frac{\partial h(a, b, c)}{\partial b}\Big|_{\overline{Y},\overline{X}} \quad \frac{\partial h(a, b, c)}{\partial c}\Big|_{\overline{Y},\overline{X}}\right]$$
$$\Sigma = \begin{bmatrix} V(\bar{y}_s) & cov(\bar{y}_s, \bar{x}) & cov(\bar{y}_s, \bar{x}_s) \\ cov(\bar{x}, \bar{y}_s) & V(\bar{x}) & cov(\bar{x}, \bar{x}_s) \\ cov(\bar{x}_s, \bar{y}_s) & cov(\bar{x}_s, \bar{x}) & V(\bar{x}_s) \end{bmatrix} \tag{35}$$

(see Wolter, 1985). Here, $h(a, b, c) = h(\bar{y}_s, \bar{x}, \bar{x}_s) = \bar{y}_{pr3}$. According to this definition, we obtain $\boldsymbol{d}$ for the proposed estimator as follows:

$$\boldsymbol{d} = [1 \quad -(R + B) \quad R + B]. \tag{36}$$

Substituting (15)–(17), (24), (25), (31) in (35) and then (35)–(36) in (34), we get (33). Hence, the theorem. □

3. Efficiency Comparisons

In this section, we compare the MSE of proposed estimators with the MSE of sample mean and ratio-type estimators in the case of missing data. From these comparisons, we obtain the conditions where the proposed estimators are more efficient than mentioned estimators.

Theorem 3.1. *All proposed estimators, given in* (11), (21), *and* (28), *are more efficient than the sample mean estimator, given in* (1), *when the following condition is satisfied*:

$$R^2 < B^2. \tag{37}$$

Proof. Comparing (19), (26), and (33) with (2), respectively, it is easily shown that we get the condition (37). □

Theorem 3.2. *First proposed estimator, given in* (11), *is more efficient than the ratio-type estimator, given in* (3), *when the following condition is satisfied*:

$$\left(\frac{1}{n}-\frac{1}{N}\right)S_x^2(R^2-B^2)-\left(\frac{1}{r}-\frac{1}{n}\right)(R^2S_x^2-2RS_{xy})<0. \tag{38}$$

Proof. Comparing (19) with (4), we easily get the condition (38). □

Theorem 3.3. *Second proposed estimator, given in* (21), *is more efficient than the ratio-type estimator, given in* (3), *when the following condition is satisfied*:

$$\left(\frac{1}{r}-\frac{1}{N}\right)S_x^2(R^2-B^2)-\left(\frac{1}{r}-\frac{1}{n}\right)(R^2S_x^2-2RS_{xy})<0. \tag{39}$$

Proof. Comparing (26) with (4), we easily get the condition (39). □

Theorem 3.4. *Third proposed estimator, given in* (28), *is more efficient than the ratio-type estimator, given in* (3), *when the following condition is satisfied*:

$$S_{xy}(2R-B)<0. \tag{40}$$

Proof. Comparing (33) with (4), we easily get the condition (40). □

Theorem 3.5. *First proposed estimator, given in* (11), *is more efficient than the estimators in Singh and Horn* (2000) *and Singh and Deo* (2003), *given in* (6) *and* (7), *respectively, when the following condition is satisfied*:

$$\left(\frac{1}{n}-\frac{1}{N}\right)R^2+\left(\frac{1}{r}+\frac{1}{N}-\frac{2}{n}\right)B^2<0. \tag{41}$$

Proof. Comparing (19) with (9) or (A1), we obtain:

$$\left(\frac{1}{n}-\frac{1}{N}\right)S_x^2(R^2-B^2)<\left(\frac{1}{r}-\frac{1}{n}\right)\left[R^2S_x^2-2RS_{yx}-S_x^2(B-R)^2\right]$$
$$\left(\frac{1}{n}-\frac{1}{N}\right)S_x^2(R^2-B^2)<\left(\frac{1}{r}-\frac{1}{n}\right)\left[-2RS_{yx}-S_x^2B^2+2BRS_x^2\right]$$
$$\left(\frac{1}{n}-\frac{1}{N}\right)S_x^2(R^2-B^2)<\left(\frac{1}{r}-\frac{1}{n}\right)\left[BS_x^2(2R-B)-2RBS_x^2\right]$$
$$\left(\frac{1}{n}-\frac{1}{N}\right)(R^2-B^2)-\left(\frac{1}{r}-\frac{1}{n}\right)\left[B(2R-B)-2RB\right]<0$$

$$\left(\frac{1}{n}-\frac{1}{N}\right)(R^2-B^2)+\left(\frac{1}{r}-\frac{1}{n}\right)B^2<0$$

$$\left(\frac{1}{n}-\frac{1}{N}\right)R^2+\left(\frac{1}{r}+\frac{1}{N}-\frac{2}{n}\right)B^2<0.$$

By this way, we get the condition (41). □

Theorem 3.6. *Second proposed estimator, given in* (21), *is more efficient than the estimators in Singh and Horn* (2000) *and Singh and Deo* (2003), *given in* (6) *and* (7), *respectively, when the following condition is satisfied*:

$$\left(\frac{1}{r}-\frac{1}{N}\right)R^2-\left(\frac{1}{n}-\frac{1}{N}\right)B^2<0. \tag{42}$$

Proof. Comparing (26) with (9) or (A1), similar with the proof in Theorem 3.5, we can easily get the condition (42). □

Theorem 3.7. *Third proposed estimator, given in* (28), *is more efficient than the estimators in Singh and Horn* (2000) *and Singh and Deo* (2003), *given in* (6) *and* (7), *respectively, when the following condition is satisfied*:

$$R^2<0. \tag{43}$$

However, this condition is never satisfied so the estimators in Singh and Horn (2000) *and Singh and Deo* (2003) *are always more efficient than the third proposed estimator.*

Proof. Comparing (33) with (9) or (A1), similar with the proof in Theorem 3.5 we can easily get the condition (43). □

4. Numerical Illustration

We have used the data of Kadilar and Cingi (2003) in this section. However, we have considered the data of all 19 counties (Ayvalik, Bandirma, Edremit, Erdek, etc.) of only 1 city (Balikesir) in Turkey in 1999, as we are interested in simple random sampling here. These data, concerning the level of apple production as the study variable and number of apple trees as the auxiliary variable, are used to compare efficiencies of proposed estimators with sample mean, ratio-type estimators, and the estimators in Singh and Horn (2000) and Singh and Deo (2003) numerically.

In Table 1, we observe the statistics about the population. Note that we take the sample of size as $n=10$ from $N=19$ by SRSWOR method and let $r=8$. We would like to remark that the coefficient of correlation between the auxiliary and study variables is 0.88 so this data set is suitable for the estimators using the ratio method.

If we performed an empirical study, we would do the following steps.

Step 1. We select all possible $M=\binom{19}{10}=92{,}378$ samples from the population.

Step 2. We drop two units randomly from each sample corresponding to the study variable.

Table 1
Data statistics

$N = 19$	$\overline{Y} = 575.00$	$S_{yx} = 9738388.33$
$n = 10$	$\overline{X} = 13573.68$	$R = 0.04$
$r = 8$	$S_y = 858.36$	$B = 0.06$
$\rho = 0.88$	$S_x = 12945.38$	

Step 3. For each sample, we estimate the population mean using mean method in (1), ratio method in (3), first proposed estimator in (11), second proposed estimator in (21), and third proposed estimator in (28).

Step 4. We compute exact MSE values of each method using the following equation:

$$MSE(\bar{y}_i) = \frac{1}{M}\sum_{j=1}^{M}(\bar{y}_{ij} - \overline{Y})^2, \quad i = s, r, pr1, pr2, pr3. \tag{44}$$

Instead of using these steps, we can compute approximate values of MSE for each estimator using (2), (4), (8), (9), (19), (26), and (33). These MSE values, which are given in Table 2, are expected to be nearly the same as the values computed from (44). However, the results of simulation would be less reliable than those of theory, because of the availability of many possible samples.

When we examine Table 2, we observe that second proposed estimator has the smallest MSE value among all estimators. It is an expected result, because the condition (39) and (42) are satisfied. The values of the expressions in (39) and (42) are −6082.32 and −3.01E-05, respectively. However, the conditions (38), (40), and (41) are not satisfied as the values of expressions in these conditions are positive. All proposed estimators have smaller MSE than mean method, because the condition (37) is satisfied for this data set.

From the result of this numerical illustration, we infer that second proposed estimator is the most efficient estimator for this data set. It is worth to point out that this inference changes depending on values of r and n.

Table 2
MSE values of estimators

$\bar{y}_s$	53319.89
$\bar{y}_r$	40211.43
$\bar{y}_{SH}$	39172.22
$\bar{y}_{SD}$	39172.22
$\bar{y}_{pr1}$	40758.65
$\bar{y}_{pr2}$	34129.11
$\bar{y}_{pr3}$	46690.35

5. Conclusion

We have developed estimators adapting the estimator considered in Kadilar and Cingi (2004) to the case of non response in sample survey and obtained the bias and MSE equations of the proposed estimators. Theoretically, we have demonstrated that the proposed estimators are more efficient than traditional estimators and the estimators in Singh and Horn (2000) and Singh and Deo (2003) when the conditions (37)–(42) are satisfied. In addition, we support this theoretical result numerically. We would like to mention that some other estimators can also be derived in the forthcoming studies by adding the auxiliary information, such as the population coefficient of variation or kurtosis of the auxiliary variable, to the proposed estimators given here, as in the studies of Sisodia and Dwivedi (1981) and Upadhyaya and Singh (1999).

Appendix

$$
\begin{aligned}
MSE_{\min}(\bar{y}_{SH}) &\cong MSE(\bar{y}_r) - \left(\frac{1}{r}-\frac{1}{n}\right)\left(1-\rho\frac{C_y}{C_x}\right)^2 \overline{Y}^2 C_x^2 \\
&= MSE(\bar{y}_r) - \left(\frac{1}{r}-\frac{1}{n}\right)\left(1-\frac{S_{xy}}{S_xS_y}\frac{S_y/\overline{Y}}{S_x/\overline{X}}\right)^2 \overline{Y}^2 \frac{S_x^2}{\overline{X}^2} \\
&= MSE(\bar{y}_r) - \left(\frac{1}{r}-\frac{1}{n}\right)\left(1-\frac{S_{xy}}{S_xS_y}\frac{S_y}{S_x}\frac{\overline{X}}{\overline{Y}}\right)^2 \overline{Y}^2 \frac{S_x^2}{\overline{X}^2} \\
&= MSE(\bar{y}_r) - \left(\frac{1}{r}-\frac{1}{n}\right)\left(1-\frac{B}{R}\right)^2 R^2 S_x^2 \\
&= MSE(\bar{y}_r) - \left(\frac{1}{r}-\frac{1}{n}\right)\frac{(R-B)^2}{R^2} R^2 S_x^2 \\
&= MSE(\bar{y}_r) - \left(\frac{1}{r}-\frac{1}{n}\right) S_x^2 (B-R)^2 = MSE_{\min}(\bar{y}_{SD}) \qquad \text{(A.1)}
\end{aligned}
$$

References

Cingi, H. (1994). *Sampling Theory*. Ankara, Turkey: Hacettepe University Press.

Cochran, W. G. (1977). *Sampling Techniques*. New York: John Wiley and Sons.

Kadilar, C., Cingi, H. (2003). Ratio estimators in stratified random sampling. *Biometrical J.* 45:218–225.

Kadilar, C., Cingi, H. (2004). Ratio estimators in simple random sampling. *Appl. Math. Computat.* 151:893–902.

Kadilar, C., Cingi, H. (2005). A new estimator using two auxiliary variables. *Appl. Math. Computat.* 162:901–908.

Ray, S. K., Singh, R. K. (1981). Difference-cum-ratio type estimators. *J. Ind. Statist. Assoc.* 19:147–151.

Rueda, M., Gonzalez, S. (2004). Missing data and auxiliary information in surveys. *Computat. Statist.* 19:551–567.

Rueda, M., Gonzalez, S., Arcos, A. (2005). Indirect methods of imputation of missing data based on available units. *Appl. Math. Computat.* 164:249–261.

Singh, S., Horn, S. (2000). Compromised imputation in survey sampling. *Metrika* 51:267–276.

Singh, S., Deo, B. (2003). Imputation by power transformation. *Statist. Pap.* 44:555–579.
Sisodia, B. V. S., Dwivedi, V. K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. *J. Ind. Soc. Agricult. Statist.* 33:13–18.
Upadhyaya, L. N., Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical J.* 41:627–636.
Wolter, K. M. (1985). *Introduction to Variance Estimation*. Springer-Verlag.