Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666
Vol. 8, Issue 1 (2015): 135-146

ESTIMATION OF POPULATION MEAN USING AUXILIARY INFORMATION IN NON-RESPONSE

R. Karan Singh and Nazia Naqvi*
Department of Statistics, Lucknow University, Lucknow, India
*Babasaheb Bhimrao Ambedkar University, Lucknow, India
E Mail: *naqvi.nazia8@gmail.com

Received September 27, 2014
Modified May 14, 2015
Accepted May 31, 2015

Abstract

In the presence of non-response, a class of estimators of finite population mean of study variable using the knowledge of population mean of an auxiliary variable is proposed and its bias and mean square error are found. A sub class of optimum estimators in the sense of having minimum mean square error is found and enhancing the practical utility, a sub class of estimators depending on estimated optimum value based on sample observations is also investigated in the presence of non-response.

Key Words: Double Sampling, Non-Response, Bias, Mean Square Error, Efficiency.

1. Introduction

Let y be the study variable of interest with its population mean $\overline{Y}$ and x be the auxiliary variable with known population mean $\overline{X}$. For the case of non-response in sample survey, the procedure of sub sampling the non-respondents was suggested by Hansen and Hurwitz (1946), see also Khare and Srivastava (1997). Let n be the size of the simple random sample without replacement drawn from the population of size N where n_1 of selected n units respond and n_2 sample units do not respond.

From the n_2 non-response units, $r\left(=\frac{n_2}{k}, k>1\right)$ units are selected by making extra efforts and thus giving n_1+r observations on the y character in place of n. For $\overline{y}_1$ being the sample mean based on n_1 units and $\overline{y}_2'$ being the sample mean based on r units, Hansen and Hurwitz (1946) using n_1+r observations on the y character gave the unbiased estimator of population mean given by $\overline{y}^*$ as

$$\overline{y}^* = \frac{n_1}{n}\overline{y}_1 + \frac{n_2}{n}\overline{y}_2' \tag{1.1}$$

whose variance is

$$Var\left(\overline{y}^*\right) = \frac{(1-f)}{n}S_y^2 + \frac{W_2(k-1)}{n}S_{y_2}^2 \tag{1.2}$$

where $f = \frac{n}{N}$, $W_i = \frac{N_i}{N}$, (i=1,2) and S_y^2 and $S_{y_2}^2$ are the variances for the whole population and for the non-response group of the population respectively.
With $\bar{X}$ being known and incomplete information on y and x, Rao (1986) proposed the conventional ratio estimator given by

$$T^* = \frac{\bar{y}^*}{\bar{x}^*}\bar{X} \tag{1.3}$$

where $\bar{x}^* = \frac{n_1}{n}\bar{x}_1 + \frac{n_2}{n}\bar{x}_2'$ with $\bar{x}_1$ and $\bar{x}_2'$ being the sample means based on n_1 and r observations on x respectively. Further, when $\bar{X}$ known and incomplete information on y and x, Khare and Srivastava (1997) proposed the transformed ratio type estimator given by

$$t_1 = \frac{\bar{y}^*\left(\bar{X}+L\right)}{\left(\bar{x}^*+L\right)} \tag{1.4}$$

where L is a positive constant.
Also, some more ratio type estimators may be considered as

$$t_2 = \frac{\bar{y}^*\left(\bar{x}^*+L\right)}{\left(\bar{X}+L\right)} \tag{1.5}$$

$$t_3 = \bar{y}^*\left(\frac{\bar{x}^*}{\bar{X}}\right)^k \tag{1.6}$$

$$t_4 = \bar{y}^* e^{k\left(\frac{\bar{x}^*}{\bar{X}}-1\right)} \tag{1.7}$$

$$t_5 = \bar{y}^*\left\{1+k\frac{\left(\bar{x}^*-\bar{X}\right)}{\bar{X}}\right\} \tag{1.8}$$

$$t_6 = \bar{y}^* + k\left(\bar{x}^*-\bar{X}\right) \tag{1.9}$$

$$t_7 = \bar{y}^* + \left\{\left(\frac{\bar{x}^*}{\bar{X}}\right)^k - 1\right\} \tag{1.10}$$

Seeing the forms of the estimators from (1.3) to (1.10), we define a generalized class of estimators as a function $\bar{y}_g^* = g\left(\bar{y}^*, \bar{x}^*\right)$ satisfying the validity conditions of Taylor's series expansion such that

(i) $g\left(\bar{Y}, \bar{X}\right) = \bar{Y}$

(ii) $g_1 = \left.\frac{\partial g\left(\bar{y}^*, \bar{x}^*\right)}{\partial \bar{y}^*}\right]_{P=\left(\bar{Y}, \bar{X}\right)} = 1$

$$(iii)\; g_0 = \left.\frac{\partial g\left(\bar{y}^*, \bar{x}^*\right)}{\partial \bar{x}^*}\right]_{P=\left(\bar{Y}, \bar{X}\right)} \tag{1.11}$$

It may be mentioned here that all estimators listed from (1.3) to (1.10) belong to the generalized class $\bar{y}_g^*$ of estimators; hence, their results may be easily obtained as the special cases directly from the generalized class $\bar{y}_g^*$ of estimators.

2. Bias and mean square error

Let us define,

$e_0^* = \left(\bar{y}^* - \bar{Y}\right)$, $e_1^* = \left(\bar{x}^* - \bar{X}\right)$ so that

$E\left(e_0^*\right) = E\left(e_1^*\right) = 0$ and

$$E\left(e_0^{*2}\right) = \frac{(N-n)}{Nn} S_y^2 + \frac{(k-1)N_2}{Nn} S_{y_2}^2 ,$$

$$E\left(e_1^{*2}\right) = \frac{(N-n)}{Nn} S_x^2 + \frac{(k-1)N_2}{Nn} S_{x_2}^2 ,$$

$$E\left(e_0^* e_1^*\right) = \frac{(N-n)}{Nn} S_{yx} + \frac{(k-1)N_2}{Nn} S_{yx_{(2)}}$$

where S_x^2 and $S_{x_2}^2$ are the variances of x for the whole population and for the non-response group of the population respectively, and S_{yx} and $S_{yx_{(2)}}$ are respectively the covariances between y and x for the whole population and for the non-response group of the population.

Expanding $\bar{y}_g^* = g\left(\bar{y}^*, \bar{x}^*\right)$ about the point $P = \left(\bar{Y}, \bar{X}\right)$ in third order Taylor's series, we have

$$\bar{y}_g^* = g\left(\bar{Y},\bar{X}\right) + \left(\bar{y}^* - \bar{Y}\right) g_1 + \left(\bar{x}^* - \bar{X}\right) g_0 + \frac{1}{2!}\left\{\left(\bar{y}^* - \bar{Y}\right)^2 g_{11} + \left(\bar{x}^* - \bar{X}\right)^2 g_{00} + 2\left(\bar{y}^* - \bar{Y}\right)\left(\bar{x}^* - \bar{X}\right) g_{10}\right\}$$

$$+ \frac{1}{3!}\left\{\left(\bar{y}^* - \bar{Y}\right)\frac{\partial}{\partial \bar{y}^*} + \left(\bar{x}^* - \bar{X}\right)\frac{\partial}{\partial \bar{x}^*}\right\}^3 g\left(\bar{y}_*^*, \bar{x}_*^*\right) \tag{2.1}$$

where $g_1 = \left.\frac{\partial g\left(\bar{y}^*, \bar{x}^*\right)}{\partial \bar{y}^*}\right]_{P=\left(\bar{Y}, \bar{X}\right)} = 1$, $g_0 = \left.\frac{\partial g\left(\bar{y}^*, \bar{x}^*\right)}{\partial \bar{x}^*}\right]_{P=\left(\bar{Y}, \bar{X}\right)}$,

$$g_{11} = \left.\frac{\partial^2 g\left(\bar{y}^*, \bar{x}^*\right)}{\partial \bar{y}^{*2}}\right]_{P=\left(\bar{Y}, \bar{X}\right)} = 0 ,$$

$$g_{00} = \left.\frac{\partial^2 g\left(\bar{y}^*, \bar{x}^*\right)}{\partial \bar{x}^{*2}}\right]_{P=\left(\bar{Y},\bar{X}\right)}, \quad g_{10} = \left.\frac{\partial^2 g\left(\bar{y}^*, \bar{x}^*\right)}{\partial \bar{y}^* \partial \bar{x}^*}\right]_{P=\left(\bar{Y},\bar{X}\right)} \text{ and}$$

$\bar{y}^*_* = \bar{Y} + \theta\left(\bar{y}^* - \bar{Y}\right)$, $\bar{x}^*_* = \bar{X} + \theta\left(\bar{x}^* - \bar{X}\right)$ for $0 < \theta < 1$.

On substituting $g\left(\bar{Y}, \bar{X}\right) = \bar{Y}$ and the values of the derivatives in (2.1), we have

$$\bar{y}^*_g = \bar{Y} + \left(\bar{y}^* - \bar{Y}\right)(1) + \left(\bar{x}^* - \bar{X}\right)g_0 + \frac{1}{2!}\left\{\left(\bar{y}^* - \bar{Y}_1\right)^2 .0 + \left(\bar{x}^* - \bar{X}\right)^2 g_{00} + 2\left(\bar{y}^* - \bar{Y}\right)\left(\bar{x}^* - \bar{X}\right)g_{10}\right\}$$

$$+\frac{1}{3!}\left\{\left(\bar{y}^* - \bar{Y}\right)\frac{\partial}{\partial \bar{y}^*} + \left(\bar{x}^* - \bar{X}\right)\frac{\partial}{\partial \bar{x}^*}\right\}^3 g\left(\bar{y}^*_*, \bar{x}^*_*\right)$$

or

$$\bar{y}^*_g - \bar{Y} = e^*_0 + e^*_1 g_0 + \frac{1}{2!}\left\{e^{*2}_1 g_{00} + 2e^*_0 e^*_1 g_{10}\right\} + \frac{1}{3!}\left\{\left(\bar{y}^* - \bar{Y}\right)\frac{\partial}{\partial \bar{y}^*} + \left(\bar{x}^* - \bar{X}\right)\frac{\partial}{\partial \bar{x}^*}\right\}^3 g\left(\bar{y}^*_*, \bar{x}^*_*\right) \tag{2.2}$$

Taking expectation on both sides of (2.2) and ignoring terms in e_i's (i=0,1) greater than two, the bias of $\bar{y}^*_g$ to the first degree of approximation {that is, up to terms of order $O\left(\frac{1}{n}\right)$} is

$$E\left(\bar{y}^*_g\right) - \bar{Y} = E\left(e^*_0\right) + E\left(e^*_1\right)g_0 + \frac{1}{2!}\left\{\left\{E\left(e^{*2}_1\right)g_{00} + 2E\left(e^*_0 e^*_1\right)g_{10}\right\}\right.$$

$$= \frac{1}{2}\left\{\frac{(N-n)}{Nn}S^2_x + \frac{(k-1)N_2}{Nn}S^2_{x_2}\right\}g_{00} + \left\{\frac{(N-n)}{Nn}S_{yx} + \frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}g_{10}$$

$$\text{or } Bias(\bar{y}^*_g) = \frac{(N-n)}{Nn}\left\{\frac{S^2_x}{2}g_{00} + S_{yx}g_{10}\right\} + \frac{(k-1)N_2}{Nn}\left\{\frac{S^2_{x_2}}{2}g_{00} + S_{yx_{(2)}}g_{10}\right\} \tag{2.3}$$

which shows that bias of $\bar{y}^*_g$ is of order $O\left(\frac{1}{n}\right)$; hence, for sufficiently large value of n, the bias is negligible.

Squaring both the sides of (2.2), taking expectation and ignoring terms in e_i's (i=0,1) greater than two, the mean square error of $\bar{y}^*_g$ to the first degree of approximation {that is, up to terms of order $O\left(\frac{1}{n}\right)$} is

$$E(\bar{y}^*_g - \bar{Y})^2 = E\left[e^*_0 + e^*_1 g_0\right]^2$$

$$= E\left(e^{*2}_0\right) + E\left(e^{*2}_1\right)g^2_0 + 2E\left(e^*_0 e^*_1\right)g_0$$

$$= \frac{(N-n)}{Nn}S_y^2 + \frac{(k-1)N_2}{Nn}S_{y_2}^2 + \left[\frac{(N-n)}{Nn}S_x^2 + \frac{(k-1)N_2}{Nn}S_{x_2}^2\right]g_0^2$$

$$+2\left\{\frac{(N-n)}{Nn}S_{yx} + \frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}g_0$$

$$MSE\left(\bar{y}_g^*\right) = MSE(\bar{y}^*) + \left[\frac{(N-n)}{Nn}S_x^2 + \frac{(k-1)N_2}{Nn}S_{x_2}^2\right]g_0^2 + 2\left\{\frac{(N-n)}{Nn}S_{yx} + \frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}g_0 \tag{2.4}$$

The optimum value of g_0 minimizing the mean square error of $\bar{y}_g^*$ is

$$g_0^* = -\frac{\frac{(1-f)}{n}S_{yx} + \frac{(k-1)N_2}{Nn}S_{yx_{(2)}}}{\frac{(1-f)}{n}S_x^2 + \frac{(k-1)N_2}{Nn}S_{x_2}^2} = G \tag{2.5}$$

where $f = \frac{n}{N}$ and the minimum mean square error of $\bar{y}_g^*$ is

$$MSE(\bar{y}_g^*)_{\min} = MSE\left(\bar{y}^*\right) + \frac{\left\{\frac{(1-f)}{n}S_{yx} + \frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}^2}{\frac{(1-f)}{n}S_x^2 + \frac{(k-1)N_2}{Nn}S_{x_2}^2} - 2\frac{\left\{\frac{(1-f)}{n}S_{yx} + \frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}^2}{\frac{(1-f)}{n}S_x^2 + \frac{(k-1)N_2}{Nn}S_{x_2}^2}$$

$$= MSE\left(\bar{y}^*\right) - \frac{\left\{\frac{(1-f)}{n}S_{yx} + \frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}^2}{\frac{(1-f)}{n}S_x^2 + \frac{(k-1)N_2}{Nn}S_{x_2}^2} \tag{2.6}$$

The optimum value of g_0^* in (2.5) contains some unknown parameters and lacks its practical utility to attain the minimum mean square error of $\bar{y}_g^*$ in (2.6); hence, the alternative is to replace g_0^* by its consistent estimated optimum value $\hat{g}_0^*$ based on sample observations.

From Rao (1990), the unbiased estimators of S_{yx}, S_x^2 , $S_{yx_{(2)}}$ and $S_{x_2}^2$ are respectively given by

$$\hat{S}_{yx} = \frac{1}{(n-1)}\left[(n_1-1)s_{yx_1} + \left\{\frac{n_2}{r}(r-1) + \frac{(k-1)W_2}{n}\right\}s'_{yx_2} + nW_1W_2\left(\bar{x}_1 - \bar{x}'_2\right)\left(\bar{y}_1 - \bar{y}'_2\right)\right]$$

$$\hat{S}_x^2 = \frac{1}{(n-1)}\left[(n_1-1)s_{x_1}^2 + \left\{\frac{n_2}{r}(r-1) + \frac{(k-1)W_2}{n}\right\}s'^2_{x_2} + nW_1W_2\left(\bar{x}_1 - \bar{x}'_2\right)^2\right]$$

$$s_{yx_1} = \frac{1}{(n_1-1)}\sum_{i=1}^{n_1}\left(x_i - \bar{x}_1\right)\left(y_i - \bar{y}_1\right)$$

$$\hat{S}_{yx_{(2)}} = s'_{yx_2} = \frac{1}{(r-1)}\sum_{i=1}^{r}\left(x_{i2} - \bar{x}'_2\right)\left(y_{i2} - \bar{y}'_2\right)$$

$$s^2_{x_1} = \frac{1}{(n_1-1)}\sum_{i=1}^{n_1}\left(x_i - \bar{x}_1\right)^2$$

$$\hat{S}'^2_{x_2} = s'^2_{x_2} = \frac{1}{(r-1)}\sum_{i=1}^{r}\left(x_{i2} - \bar{x}'_2\right)^2$$

which when substituted in g_0^* in (2.5) gives the estimated optimum value

$$\hat{g}_0^* = -\frac{\frac{(1-f)}{n}\hat{S}_{yx} + \frac{(k-1)N_2}{Nn}\hat{S}_{yx_{(2)}}}{\frac{(1-f)}{n}\hat{S}_x^2 + \frac{(k-1)N_2}{Nn}\hat{S}^2_{x_2}} = \hat{G} \qquad (2.7)$$

where x_{i2} and y_{i2} denote the observations on the ith unit of the sub-sampling units selected from n_2 non-respondent units.

To attain the minimum mean square error of $\bar{y}_g^*$ in (2.6), the function $\bar{y}_g^* = g\left(\bar{y}^*, \bar{x}^*\right)$ should not only involve $\left(\bar{y}^*, \bar{x}^*\right)$ satisfying (i) to (iii) in (1.11) but also $g_0^* = G$ in (2.5), but $g_0^* = G$ is unknown depending on parameters; hence, we should have the estimator as a function $\bar{y}_{ge}^* = g\left(\bar{y}^*, \bar{x}^*, \hat{G}\right)$ depending on estimated optimum $\hat{g}_0^* = \hat{G}$ in (2.7) and find the conditions which make the estimator $\bar{y}_{ge}^* = g\left(\bar{y}^*, \bar{x}^*, \hat{G}\right)$ having its mean square error to be equal to the minimum mean square error of $\bar{y}_g^*$ in (2.6).

Now expanding $\bar{y}_{ge}^* = g\left(\bar{y}^*, \bar{x}^*, \hat{G}\right)$ about the point $T = \left(\bar{Y}, \bar{X}, G\right)$ in Taylor's series, we have

$$\bar{y}_{ge}^* = g\left(\bar{Y}, \bar{X}, G\right) + \left(\bar{y}^* - \bar{Y}\right)g_{1e} + \left(\bar{x}^* - \bar{X}\right)g_{0e} + \left(\hat{G} - G\right)g_{2e} + \frac{1}{2!}\left\{\left(\bar{y}^* - \bar{Y}\right)^2 g_{11e} + \left(\bar{x}^* - \bar{X}\right)^2 g_{00e} + \left(\hat{G} - G\right)^2 g_{22e}\right.$$
$$\left. + 2\left(\bar{y}^* - \bar{Y}\right)\left(\bar{x}^* - \bar{X}\right)g_{10e} + 2\left(\bar{y}^* - \bar{Y}\right)\left(\hat{G} - G\right)g_{12e} + 2\left(\bar{x}^* - \bar{X}\right)\left(\hat{G} - G\right)g_{02e}\right\} + \ldots \qquad (2.8)$$

Similar to $\bar{y}_g^*$ from (i) to (iii) and (2.5), substituting in (2.8), we have

$$\bar{y}_{ge}^* = \bar{Y} + \left(\bar{y}^* - \bar{Y}\right).1 + \left(\bar{x}^* - \bar{X}\right)G + \left(\hat{G} - G\right)g_{2e} + \frac{1}{2!}\left\{\left(\bar{y}^* - \bar{Y}\right)^2 .0 + \left(\bar{x}^* - \bar{X}\right)^2 g_{00e} + \left(\hat{G} - G\right)^2 g_{22e}\right.$$
$$\left. + 2\left(\bar{y}^* - \bar{Y}\right)\left(\bar{x}^* - \bar{X}\right)g_{10e} + 2\left(\bar{y}^* - \bar{Y}\right)\left(\hat{G} - G\right)g_{12e} + 2\left(\bar{x}^* - \bar{X}\right)\left(\hat{G} - G\right)g_{02e}\right\} + \ldots$$

or

$$\left(\bar{y}_{ge}^{*}-\bar{Y}\right)=e_{0}^{*}+e_{1}^{*}G+e_{2}^{*}g_{2e}+\frac{1}{2!}\left\{e_{1}^{*2}g_{00e}+e_{2}^{*2}g_{22e}+2e_{0}^{*}e_{1}^{*}g_{10e}+2e_{1}^{*}e_{2}^{*}g_{12e}+2e_{0}^{*}e_{2}^{*}g_{20e}\right\}+\ldots \tag{2.9}$$

where $e_{2}^{*}=\left(\hat{G}-G\right)$, $g\left(\bar{Y},\bar{X},G\right)=\bar{Y}$,

$$g_{1e}=\left.\frac{\partial g\left(\bar{y}^{*},\bar{x}^{*},\hat{G}\right)}{\partial \bar{y}^{*}}\right]_{T}=1\ ,\ g_{0e}=\left.\frac{\partial g\left(\bar{y}^{*},\bar{x}^{*},\hat{G}\right)}{\partial \bar{x}^{*}}\right]_{T}=G\ ,$$

$$g_{2e}=\left.\frac{\partial g\left(\bar{y}^{*},\bar{x}^{*},\hat{G}\right)}{\partial \hat{G}}\right]_{T}\ ,\ g_{11e}=\left.\frac{\partial^{2} g\left(\bar{y}^{*},\bar{x}^{*},\hat{G}\right)}{\partial \bar{y}^{*2}}\right]_{T}=0\ ,$$

$$g_{00e}=\left.\frac{\partial^{2} g\left(\bar{y}^{*},\bar{x}^{*},\hat{G}\right)}{\partial \bar{x}^{*2}}\right]_{T}\ ,\ g_{22e}=\left.\frac{\partial^{2} g\left(\bar{y}^{*},\bar{x}^{*},\hat{G}\right)}{\partial \hat{G}^{2}}\right]_{T}\ ,$$

$$g_{10e}=\left.\frac{\partial^{2} g\left(\bar{y}^{*},\bar{x}^{*},\hat{G}\right)}{\partial \bar{y}^{*}\partial \bar{x}^{*}}\right]_{T}\ ,\ g_{12e}=\left.\frac{\partial^{2} g\left(\bar{y}^{*},\bar{x}^{*},\hat{G}\right)}{\partial \bar{y}^{*}\partial \hat{G}}\right]_{T}\ ,$$

$$g_{02e}=\left.\frac{\partial^{2} g\left(\bar{y}^{*},\bar{x}^{*},\hat{G}\right)}{\partial \bar{x}^{*}\partial \hat{G}}\right]_{T}\ .$$

Taking expectation on both sides of (2.9), and ignoring terms in e_i's greater than two, the bias of $\bar{y}_{ge}^{*}$ to the first degree of approximation is of order $O\left(\frac{1}{n}\right)$.

The mean square error of $\bar{y}_{ge}^{*}$ given by $E\left(\bar{y}_{ge}^{*}-\bar{Y}\right)^{2}$ to the first degree of approximation is

$$MSE\left(\bar{y}_{ge}^{*}\right)=E\left\{e_{0}^{*}+e_{1}^{*}G+e_{2}^{*}g_{2e}\right\}^{2}$$

$$=E\left\{\left(e_{0}^{*}+e_{1}^{*}G\right)^{2}+e_{2}^{*2}g_{2e}^{2}+2\left(e_{0}^{*}+e_{1}^{*}G\right)e_{2}^{*}g_{2e}\right\}$$

For $g_{2e}=0$, $MSE(\bar{y}_{ge}^{*})$ becomes

$$MSE\left(\bar{y}_{ge}^{*}\right)=E\left\{e_{0}^{*}+e_{1}^{*}G\right\}^{2}$$

$$=E\left\{e_{0}^{*2}+2e_{0}^{*}e_{1}^{*}G+e_{1}^{*2}G^{2}\right\}$$

$$=MSE\left(\bar{y}^{*}\right)+2\left\{\frac{(1-f)}{n}S_{yx}+\frac{(k-1)N_{2}}{Nn}S_{yx_{(2)}}\right\}\left[-\frac{\left\{\frac{(1-f)}{n}S_{yx}+\frac{(k-1)N_{2}}{Nn}S_{yx_{(2)}}\right\}}{\left\{\frac{(1-f)}{n}S_{x}^{2}+\frac{(k-1)N_{2}}{Nn}S_{x_{2}}^{2}\right\}}\right]$$

$$+\frac{\left\{\frac{(1-f)}{n}S_{yx}+\frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}^2}{\left\{\frac{(1-f)}{n}S_x^2+\frac{(k-1)N_2}{Nn}S_{x_2}^2\right\}^2}\left\{\frac{(1-f)}{n}S_x^2+\frac{(k-1)N_2}{Nn}S_{x_2}^2\right\}$$

$$=MSE\left(\bar{y}^*\right)-\frac{\left\{\frac{(1-f)}{n}S_{yx}+\frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}^2}{\frac{(1-f)}{n}S_x^2+\frac{(k-1)N_2}{Nn}S_{x_2}^2} \tag{2.10}$$

which is equal to $MSE\left(\bar{y}_g^*\right)_{\min}$ in (2.6) if $g_{2e}=0$. Thus, the estimator $\bar{y}_{ge}^*=g\left(\bar{y}^*,\bar{x}^*,\hat{G}\right)$ attains the minimum mean square error of $\bar{y}_g^*$ in (2.6) if $\bar{y}_{ge}^*=g\left(\bar{y}^*,\bar{x}^*,\hat{G}\right)$ is based on estimated optimum $\hat{G}$ satisfying at the point $T=\left(\bar{Y},\bar{X},G\right)$, the following conditions

$$\left.\begin{aligned} &g\left(\bar{Y},\bar{X},G\right)=\bar{Y}\\ &g_{1e}=\left.\frac{\partial g\left(\bar{y}^*,\bar{x}^*,\hat{G}\right)}{\partial\bar{y}^*}\right]_T=1\\ &g_{0e}=\left.\frac{\partial g\left(\bar{y}^*,\bar{x}^*,\hat{G}\right)}{\partial\bar{x}^*}\right]_T=G\\ &g_{2e}=\left.\frac{\partial g\left(\bar{y}^*,\bar{x}^*,\hat{G}\right)}{\partial\hat{G}}\right]_T=0 \end{aligned}\right\} \tag{2.11}$$

3. Some particular estimators depending on estimated optimum

(a) For the estimator $t_1=\frac{\bar{y}^*\left(\bar{X}+L\right)}{\left(\bar{x}^*+L\right)}$ by Khare and Srivastava (1997), we have

$$g_0=\left.\frac{\partial\left\{\bar{y}^*\left(\bar{X}+L\right)/\left(\bar{x}^*+L\right)\right\}}{\partial\bar{x}^*}\right]_{(\bar{Y},\bar{X})}=-\frac{\bar{Y}}{\left(\bar{X}+L\right)}$$

which when substituted in (2.4), we get from the general expression for mean square error of $\bar{y}_g^*$ to be

$$MSE(t_1)=MSE\left(\bar{y}^*\right)+\left(\frac{\bar{Y}}{\bar{X}+L}\right)^2\left\{\frac{(1-f)}{n}S_x^2+\frac{(k-1)N_2}{Nn}S_{x_2}^2\right\}$$
$$-2\left(\frac{\bar{Y}}{\bar{X}+L}\right)\left\{\frac{(1-f)}{n}S_{yx}+\frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}$$

which is the same result as obtained by Khare and Srivastava (1997) as a special case of the general result in (2.4) of $\bar{y}_g^*$ for $g_0 = -\frac{\bar{Y}}{\left(\bar{X}+L\right)}$.

Also, for the optimum value in (2.5)

$$g_0^* = -\frac{\bar{Y}}{\left(\bar{X}+L\right)} = -\frac{\left\{\frac{(1-f)}{n}S_{yx} + \frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}}{\left\{\frac{(1-f)}{n}S_x^2 + \frac{(k-1)N_2}{Nn}S_{x_2}^2\right\}} = G \tag{3.1}$$

gives the same optimum value of L as given by

$$L_{opt} = \bar{X}\left[\frac{\left\{\frac{(1-f)}{n}S_x^2 + \frac{(k-1)N_2}{Nn}S_{x_2}^2\right\}R}{\left\{\frac{(1-f)}{n}S_{yx} + \frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}} - 1\right] = -\frac{\bar{Y}}{G} - \bar{X} \tag{3.2}$$

and the same minimum mean square error of t_1 as a special case of the general $MSE\left(\bar{y}_g^*\right)_{\min}$ in (2.6) to be

$$MSE(t_1) = MSE\left(\bar{y}^*\right) - \frac{\left\{\frac{(1-f)}{n}S_{yx} + \frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}^2}{\left\{\frac{(1-f)}{n}S_x^2 + \frac{(k-1)N_2}{Nn}S_{x_2}^2\right\}}$$

as obtained by Khare and Srivastava (1997). Considering the estimated optimum value $\hat{L}_{opt} = -\frac{\bar{y}^*}{\hat{G}} - \bar{X}$, we get the estimator depending on estimated optimum value $\hat{L}_{opt}$ to be

$$\begin{aligned} t_{1e} &= \frac{\bar{y}^*\left(\bar{X}+\hat{L}_{opt}\right)}{\left(\bar{x}^*+\hat{L}_{opt}\right)} \\ &= \bar{y}^*\left(\bar{X} - \frac{\bar{y}^*}{\hat{G}} - \bar{X}\right)\Big/\left(\bar{x}^* - \frac{\bar{y}^*}{\hat{G}} - \bar{X}\right) \\ &= \bar{y}^*\Big/\left\{1 - \frac{\left(\bar{x}^*-\bar{X}\right)\hat{G}}{\bar{y}^*}\right\} \end{aligned}$$

satisfying the conditions of $\bar{y}_{ge}^* = g\left(\bar{y}^*, \bar{x}^*, \hat{G}\right)$ in (2.11), since

$$t_{1e} \text{ at}\left(\bar{Y}, \bar{X}, G\right) = \bar{Y}\Big/\left\{1 - \frac{\left(\bar{X}-\bar{X}\right)\hat{G}}{\bar{Y}}\right\} = \bar{Y}$$

$$g_{1e} = \left.\frac{\partial t_{1e}}{\partial \overline{y}^*}\right]_T = \left.\frac{\left\{1 - \frac{2\left(\overline{x}^* - \overline{X}\right)}{\overline{y}^*}\hat{G}\right\}}{\left\{1 - \frac{\left(\overline{x}^* - \overline{X}\right)}{\overline{y}^*}\hat{G}\right\}^2}\right]_T = 1$$

$$g_{0e} = \left.\frac{\partial t_{1e}}{\partial \overline{x}^*}\right]_T = \left.\frac{\hat{G}}{\left\{1 - \frac{\left(\overline{x}^* - \overline{X}\right)}{\overline{y}^*}\hat{G}\right\}^2}\right]_T = G$$

$g_{2e} = \left.\frac{\partial t_{1e}}{\partial \hat{G}}\right]_T = 0$; hence, the estimator $t_{1e} = \frac{\overline{y}^*}{\left\{1 + \frac{\left(\overline{x}^* - \overline{X}\right)\hat{G}}{\overline{y}^*}\right\}}$ depending on estimated optimum value attains the minimum mean square error in (2.10) given by

$$MSE\left(\overline{y}^*\right) = \frac{\left\{\frac{(1-f)}{n}S_{yx} + \frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}^2}{\left\{\frac{(1-f)}{n}S_x^2 + \frac{(k-1)N_2}{Nn}S_{x_2}^2\right\}}.$$

(b) For the estimator $t_4 = \overline{y}^* e^{k\left(\frac{\overline{x}^*}{\overline{X}} - 1\right)}$, we have

$$g_0 = \left.\frac{\partial \overline{y}^* e^{k\left(\frac{\overline{x}^*}{\overline{X}} - 1\right)}}{\partial \overline{x}^*}\right]_{\left(\overline{Y}, \overline{X}\right)} = k\left(\frac{\overline{Y}}{\overline{X}}\right) = kR,$$

which when substituted in (2.4), we get directly the mean square error of the estimator $t_4 = \overline{y}^* e^{k\left(\frac{\overline{x}^*}{\overline{X}} - 1\right)}$ to be

$$MSE\left(\overline{y}^*\right) + k^2R^2\left\{\frac{(1-f)}{n}S_x^2 + \frac{(k-1)N_2}{Nn}S_{x_2}^2\right\} + 2kR\left\{\frac{(1-f)}{n}S_{yx} + \frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}.$$

Minimizing value of g_0 is

$$g_0^* = kR = -\frac{\left\{\frac{(1-f)}{n}S_{yx} + \frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}}{\left\{\frac{(1-f)}{n}S_x^2 + \frac{(k-1)N_2}{Nn}S_{x_2}^2\right\}} = G$$

giving $k=\dfrac{G}{R}$ which when substituted in (2.6), gives the minimum mean square error of t_4 to be

$$=MSE\left(\bar{y}^*\right)+G^2\left\{\frac{(1-f)}{n}S_x^2+\frac{(k-1)N_2}{Nn}S_{x_2}^2\right\}+2G\left\{\frac{(1-f)}{n}S_{yx}+\frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}$$

$$=MSE\left(\bar{y}^*\right)-\frac{\left\{\frac{(1-f)}{n}S_{yx}+\frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}^2}{\left\{\frac{(1-f)}{n}S_x^2+\frac{(k-1)N_2}{Nn}S_{x_2}^2\right\}}.$$

Now, considering the estimated optimum value $\hat{g}_0^*=\hat{k}\hat{R}=\hat{k}\left(\dfrac{\bar{y}^*}{\bar{x}^*}\right)=\hat{G}$ giving $\hat{k}=\dfrac{\hat{G}\bar{x}^*}{\bar{y}^*}$ which gives the estimator depending on estimated optimum value to be $t_{4e}=\bar{y}^*e^{\frac{\hat{G}\bar{x}^*}{\bar{y}^*}\left(\frac{\bar{x}^*}{\bar{X}}-1\right)}$ satisfying the conditions

(i) $\left.\bar{y}^*e^{\frac{\hat{G}\bar{x}^*}{\bar{y}^*}\left(\frac{\bar{x}^*}{\bar{X}}-1\right)}\right]_{(\bar{Y},\bar{X},G)}=\bar{Y}$

(ii) $g_{1e}=\left[1.e^{\frac{\hat{G}\bar{x}^*}{\bar{y}^*}\left(\frac{\bar{x}^*}{\bar{X}}-1\right)}+\bar{y}^*e^{\frac{\hat{G}\bar{x}^*}{\bar{y}^*}\left(\frac{\bar{x}^*}{\bar{X}}-1\right)}\left\{-\frac{\hat{G}\bar{x}^*}{\bar{y}^{*2}}\left(\frac{\bar{x}^*}{\bar{X}}-1\right)\right\}\right]_{(\bar{Y},\bar{X},G)}=1$

(iii) $g_{0e}=\left[\bar{y}^*e^{\frac{\hat{G}\bar{x}^*}{\bar{y}^*}\left(\frac{\bar{x}^*}{\bar{X}}-1\right)}\left\{\frac{\hat{G}}{\bar{y}^*}\left(\frac{\bar{x}^*}{\bar{X}}-1\right)+\frac{\hat{G}\bar{x}^*}{\bar{y}^*\bar{X}}\right\}\right]_{(\bar{Y},\bar{X},G)}=G$

and (iv) $g_{2e}=\left.\bar{y}^*e^{\frac{\hat{G}\bar{x}^*}{\bar{y}^*}\left(\frac{\bar{x}^*}{\bar{X}}-1\right)}\left\{\frac{\bar{x}^*}{\bar{y}^*}\left(\frac{\bar{x}^*}{\bar{X}}-1\right)\right\}\right]_{(\bar{Y},\bar{X},G)}=0$

of (2.11); hence, the estimator $t_{4e}=\bar{y}^*e^{\frac{\hat{G}\bar{x}^*}{\bar{y}^*}\left(\frac{\bar{x}^*}{\bar{X}}-1\right)}$ with estimated optimum $\hat{G}$ in (2.7) attains the minimum mean square error in (2.10) given by

$$MSE\left(\bar{y}^*\right)-\frac{\left\{\frac{(1-f)}{n}S_{yx}+\frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}^2}{\left\{\frac{(1-f)}{n}S_x^2+\frac{(k-1)N_2}{Nn}S_{x_2}^2\right\}}.$$

(c) For the conventional ratio estimator $T_R = \frac{\overline{y}^*}{\overline{x}^*}\overline{X}$ by Hansen and Hurwitz (1946), we have $g_0 = -\frac{\overline{Y}}{\overline{X}}$ which when substituted in (2.4), we get the minimum mean square error of T_R as a special case of $\overline{y}_g^*$ to be

$$= MSE\left(\overline{y}^*\right) + \left\{\frac{(1-f)}{n}S_x^2 + \frac{(k-1)N_2}{Nn}S_{x_2}^2\right\}\frac{\overline{Y}^2}{\overline{X}^2} - 2\left\{\frac{(1-f)}{n}S_{yx} + \frac{(k-1)N_2}{Nn}S_{yx_{(2)}}\right\}\frac{\overline{Y}}{\overline{X}}$$

$$= \frac{(1-f)}{n}S_y^2 + \frac{(1-f)}{n}R^2S_x^2 - 2\frac{(1-f)}{n}S_{yx} + \frac{(k-1)N_2}{Nn}\left\{S_{y_2}^2 + R^2S_{x_2}^2 - 2RS_{yx_{(2)}}\right\}$$

$$= \frac{(1-f)}{n}\left[S_y^2 + R^2S_x^2 - 2RS_{yx}\right] + \frac{(k-1)N_2}{Nn}\left[S_{y_2}^2 + R^2S_{x_2}^2 - 2RS_{yx_{(2)}}\right]$$

which is the same result as obtained by Hansen and Hurwitz (1946). It may be mentioned here that T_R of Hansen and Hurwitz (1946) cannot attain the minimum mean square error since T_R does not satisfy the conditions of G or $\hat{G}$.

(d) All the results of the remaining estimators t_2, t_3 and t_5 to t_7 may be easily obtained as special cases of the general results given in (2.4), (2.5), (2.6) and (2.10) as found in (a) and (b).

References

1. Cochran, W.G. (1977). Sampling Techniques, New York, John Wiley & Sons.
2. Hansen, M.H. and Hurwitz, W.N. (1946). The problem of non-response in sample surveys, Journal of American Statistical Association, 41, p. 517-529.
3. Khare, B.B. and Srivastava, S. (1997). Transformed ratio type estimators for the population mean in the presence of non-response, Communications in Statistics-Theory and Methods, 26(7), p. 1779-1791.
4. Rao, P.S.R.S. (1986). Ratio estimation with sub-sampling the non-respondents, Survey Methodology, 12(2), p. 217-230.
5. Rao, P.S.R.S. (1990). Regression estimators with sub-sampling of non-respondents, Data Quality Control: Theory and Pragmatics (Gunar e, Liepins and V.R.R. Uppuluri, Eds), Marcel Dekker, New York, p. 191-208.
6. Sisodia, B.V. and Dwivedi, V. K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable, Journal of Indian Society of Agricultural Statistics, 33, p. 13-18.
7. Upadhyaya, L.N. and Singh, H.P. (1999). Use of transformed auxiliary variable in estimating the finite population mean, Biometrical Journal, 41(5), p. 627-636.