STATISTICS IN TRANSITION new series, June 2025
Vol. 26, No. 2, pp. xxx–xxx, https://doi.org/10.59139/stattrans-2025-xxx
Received – xx.xx.xxxx; accepted – xx.xx.xxxx

Formulation of estimator for population mean in stratified successive sampling using memory-based information

Sanjoy Majumder[1], Arnab Bandyopadhyay[2], Arindam Gupta[3]

Abstract

In study described in this article, we developed a memory type estimator for the population mean in stratified successive sampling. We used the past sample information together with the current sample information through hybrid exponentially weighted moving averages statistics. We have also used the information available on auxiliary variable to construct the proposed estimator. We studied the properties of the proposed estimator. Further, we examined the performance of the proposed estimator in comparison with conventional estimator of the population mean and the results are demonstrated by using the data set of simulation as well as natural population. After observing the auspicious findings, we suggest that the proposed estimator can be applied to solve real-life problems.

Key words: successive sampling, HEWMA, regression estimator, variance, minimum variance, efficiency.

1. Introduction

In the context of socio-economic surveys, our society is composed of various classes of individuals like business owners, salaried persons, daily wages labourers, etc., and these distinct classes have various levels of income and expenditure. Therefore, when assessing these socio-economic parameters, it is needed to categorize these heterogeneous units of the population into different homogenous strata based on their socio-economic status, which necessitates the application of stratified random sampling techniques. It may also be noticed that socio-economic factors such as income and expenditure are changing over a period of time. When the characteristics are liable to change over time, successive sampling is the most preferred method for estimating

[1] Department of Mathematics and Statistics, Aliah University, Kolkata-700160, India. E-mail: sanjoybiostat@gmail.com. ORCID: https://orcid.org/0000-0001-8320-003x

[2] Department of Basic Science and Humanities (Mathematics), Dr. B. C. Roy Engineering College, Durgapur 713206, India. E-mail: arnab.bandyopadhyay4@gmail.com. ORCID: https://orcid.org/0000-0002-0769-7491.

[3] Department of Statistics, The University of Burdwan, Burdwan-713104, India. E-mail: guptaarin@gmail.com. ORCID: https://orcid.org/0000-0001-9381-5150

the population parameters at different points of time along with measuring the change over a period of time. Consequently, estimating these socio-economic parameters for various classes of individuals, stratified successive sampling may be the effective technique. Jessan (1942) was first introduced successive sampling under simple random sampling and later it was further developed by Patterson (1950), Rao and Graham (1964), Sen (1971), Sen (1972, 1973), Das (1982), Chaturvedi and Tripathi (1983) and many others. This work was further developed by, Singh and Singh (2001), Singh (2003) among others.

In many circumstances, information regarding auxiliary variable is easily available on both the first and second occasions. Using the information on auxiliary variable (first and second occasions) Feng and Zou (1997) suggested an estimator for population mean on current occasion. It was further extended by Birader and Singh (2001), Singh and Karna (2009), Singh and Vishwakarma (2009), Singh *et al.* (2011) and many others, who suggested estimation procedures for population mean on current (second) occasions in successive sampling (two occasions).

It is common practice in successive sampling that we treat the variables y and x as the character under study at current occasion and previous occasion respectively. Therefore, we estimate the parameters on y and x based on information gathered on single occasion i.e., current occasion and previous occasion respectively. It is obvious that the use of information gathered from past samples at different points of time (i.e., occasions) along with the information on the current occasion improves the performance of the estimator. It is noted that hybrid exponential weighted moving average statistics (HEWMA) help us in developing the estimator for the parameter on the variable y at the current occasion (i.e., y_t at time t) based on the information from the previous occasion such as y_{t-1} from t-1 and y_{t-2} from time t-2 and so on. It may be observed that several authors including Noor-ul-Amin (2020, 2021) and Aslam *et al.* (2020, 2023), Bhushan *et al.* (2023) developed effective estimation technique using HEWMA statistics for the population parameters in sample surveys. The utilization of HEWMA statistics may be proven to be commendable for estimating population parameter in successive sampling where collection of information at different points of time is necessary.

However, it may be noted that almost no attempt has been done for estimating population mean in stratified successive sampling using memory-based information. Motivated with these arguments, we have formulated a memory type estimator for population mean in stratified successive sampling and we examined detailed properties of the proposed estimator through empirical investigation carried over the data set of simulation as well as natural population.

2. Sample structure

Let $U = \{U_1, U_2, U_3, \ldots, U_N\}$ be finite population of size N which splits into K non overlapping strata or homogeneous strata, with each stratum containing N_h (h = 1, 2, . . ., K) units such that $N = \sum_{h=1}^{K} N_h$. The population U has been sampled over two occasions. The characters under study are denoted by x and y at the first and second (current) occasion respectively. A simple random sample S_{n_h} (WOR) of n_h units is drawn from each of the h^{th} stratum having N_h (h = 1, 2, . . ., K) units of the first occasion. A random subsample S_{m_h} (h = 1, 2, …, K) of m_h units is retained (matched) from each of the n_h units of the sample S_{n_h} (h = 1, 2, . . ., K) for its use on the second occasion. Again, a fresh simple random sample S_{u_h} (WOR) of u_h ($n_h - m_h$) units is drawn on the second occasion from each of the N_h unit of the population so that the sample size on the second occasion is also n_h ($n_h = u_h + m_h$; h = 1, 2, . . ., K).

Let y, x and z be the study variable at current occasion, study variable at previous occasion and auxiliary variable which is stable over occasions respectively taking values y_{hi}, x_{hi} and z_{hi} for the i^{th} unit (i = 1, 2, . . ., N_h) of the h^{th} stratum (h = 1, 2, . . ., K).

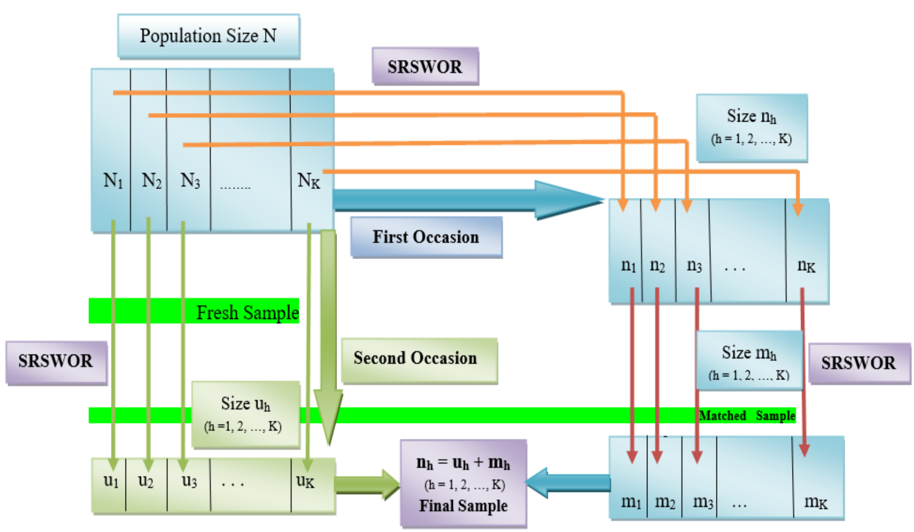

Figure 1: Sample structure for two-occasion stratified successive sampling

Henceforth, we use the following notations:

$\bar{Y}_h = \sum_{i=1}^{N_h} \frac{y_{hi}}{N_h}$, $\bar{X}_h = \sum_{i=1}^{N_h} \frac{x_{hi}}{N_h}$, $\bar{Z}_h = \sum_{i=1}^{N_h} \frac{z_{hi}}{N_h}$: Population means of the respective variables on the h^{th} stratum (h = 1, 2, . . ., K).

$W_h = \frac{N_h}{N}$: The original weight of the h^{th} stratum (h = 1, 2, . . ., K).

$\bar{Y} = \sum_{h=1}^{K} \bar{Y}_h W_h$, $\bar{X} = \sum_{h=1}^{K} \bar{X}_h W_h$, $\bar{Z} = \sum_{h=1}^{K} \bar{Z}_h W_h$: Overall population means of the respective variables.

$\bar{y}_{m_h} = \frac{1}{m_h} \sum_{i=1}^{m_h} y_{hi}$, $\bar{x}_{m_h} = \frac{1}{m_h} \sum_{i=1}^{m_h} x_{hi}$, $\bar{z}_{m_h} = \frac{1}{m_h} \sum_{i=1}^{m_h} z_{hi}$: Sample means of the respective variables based on the sample S_{m_h} of size m_h on the h^{th} stratum (h = 1, 2, . . ., K).

$\bar{x}_{n_h} = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}$, $\bar{z}_{n_h} = \frac{1}{n_h} \sum_{i=1}^{n_h} z_{hi}$: Sample means of the respective variables based on the S_{n_h} of size n_h on the h^{th} stratum (h = 1, 2, . . ., K).

$\bar{z}_{u_h} = \frac{1}{u_h} \sum_{i=1}^{u_h} z_{hi}$: Sample mean of the variable z based on the sample S_{u_h} of size u_h on the h^{th} stratum (h = 1, 2, . . ., K).

$C_{y_h} = \frac{S_{y_h}}{\bar{Y}_h}$, $C_{x_h} = \frac{S_{x_h}}{\bar{X}_h}$, $C_{z_h} = \frac{S_{z_h}}{\bar{Z}_h}$: Coefficient of variations of the respective variables based on the h^{th} stratum (h = 1, 2, . . ., K).

ρ_{yx_h}, ρ_{yz_h}, ρ_{zx_h} : Correlation coefficients between the respective variables based on the h^{th} stratum (h = 1, 2, . . ., K).

β_{yx_h}, β_{xz_h}: Sample regression coefficients between the respective variables based on the h^{th} stratum $(h = 1, 2, \ldots, K)$.

$S^2_{y_h} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(y_{hi} - \bar{Y}_h \right)^2$: Population mean square of the variable y based on the h^{th} stratum (h = 1, 2, ..., K).

$S^2_{x_h}$, $S^2_{z_h}$: Population mean squares of the variables x and z based on the h^{th} stratum (h = 1, 2, …, K).

3. Conventional hybrid exponentially weighted moving average statistics (HEWMA)

Let us assume $X_1, X_2, \ldots, X_n$ are the independently and identically distributed random variables. From this we build a sequence $HE_1, HE_2, \ldots, HE_n$ such as

$$E_t = \gamma_2 \overline{X}_t + (1-\gamma_2)E_{t-1},\ 0 < \gamma_2 \leq 1 \tag{1}$$

$$HE_t = (1-\gamma_1)HE_{t-1} + \gamma_1 E_t,\ 0 < \gamma_1 \leq 1 \tag{2}$$

where γ_1 and γ_2 are real scalars and $\overline{X}_t$ is the mean of the variable X at the time t (i.e., current occasion). It is also noticed from the above-mentioned statistic that the efficiency of the estimator can be enhanced by incorporating the information from the current sample together with the information available from the previous samples such as from time t-1, from time t-2 and so on, and HE_t denotes the hybrid exponentially weighted moving average statistic which is based upon E_t (exponentially weighted moving average). It was first introduced by Roberts (1959) in control charting. The initial values of E_t and HE_t are taken as usual mean, which may be estimated from a pilot survey; it is considered as zero i.e., $HE_0 = Et_0 = 0$. This statistic was proposed by Haq (2013). The respective expected value and variance of HEWMA statistic evaluated by Haq (2016) and and Noor-ul-Amin (2020) is given by

$$E(HE_t) = \mu \tag{3}$$

$$\mathrm{Var}(\mathrm{HE_t}) = \frac{(\gamma_1\gamma_2)^2}{(\gamma_1-\gamma_2)^2}\left[\frac{(1-\gamma_1)^2\{1-(1-\gamma_1)^{2t}\}}{1-(1-\gamma_1)^2} + \frac{(1-\gamma_2)^2\{1-(1-\gamma_2)^{2t}\}}{1-(1-\gamma_2)^2} - \frac{2(1-\gamma_1)(1-\gamma_2)\{1-(1-\gamma_1)^t(1-\gamma_2)^t\}}{1-(1-\gamma_1)(1-\gamma_2)}\right]\frac{\sigma^2}{n} \tag{4}$$

where $t \geq 1$, μ and σ^2 are the mean and variance of the variable of interest.

The limiting form of the variance is described as

$$Var(HE_t) = \frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2}\left[\frac{(1-\gamma_1)^2}{1-(1-\gamma_1)^2} + \frac{(1-\gamma_2)^2}{1-(1-\gamma_2)^2} - \frac{2(1-\gamma_1)(1-\gamma_2)}{1-(1-\gamma_1)(1-\gamma_2)}\right]\frac{\sigma^2}{n} = \frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2}R\frac{\sigma^2}{n} \tag{5}$$

where $R = \frac{(1-\gamma_1)^2}{1-(1-\gamma_1)^2} + \frac{(1-\gamma_2)^2}{1-(1-\gamma_2)^2} - \frac{2(1-\gamma_1)(1-\gamma_2)}{1-(1-\gamma_1)(1-\gamma_2)}$.

4. Proposed memory type estimator

Using the concept of HEWMA statistics for the variables x, y, z under stratified successive sampling, we have developed an estimation strategy in the following way:

a. HEWMA statistics for the variable y, based on the sample S_{m_h} of size m_h (h = 1, 2, …., K) has been defined as

$$\begin{aligned} E_{ty_{m_h}} &= \gamma_2 \overline{y}_{tm_h} + (1-\gamma_2)E_{(t-1)y_{m_h}} \\ A'_{tm_h} &= \gamma_1 E_{ty_{m_h}} + (1-\gamma_1)A'_{(t-1)m_h} \end{aligned} \tag{6}$$

where $\overline{y}_{tm_h}$ indicates sample mean of the variable y, which is based on the sample S_{m_h} at time t.

b. HEWMA statistics for the variable x, which is based on the sample S_{m_h} of size m_h (h = 1, 2, …., K) has been defined as

$$\begin{aligned} E_{(t-1)x_{m_h}} &= \gamma_2 \overline{x}_{(t-1)m_h} + (1-\gamma_2) E_{(t-2)x_{m_h}} \\ B'_{(t-1)m_h} &= \gamma_1 E_{(t-1)x_{m_h}} + (1-\gamma_1) B'_{(t-2)m_h} \end{aligned} \tag{7}$$

where $\overline{x}_{(t-1)m_h}$ indicates sample mean of the variable x, which is based on the sample S_{m_h} at time t-1.

c. HEWMA statistics for the variable x, which is based on the sample S_{n_h} of size n_h (h = 1, 2, …., K) has been defined as

$$\begin{aligned} E_{(t-1)x_{n_h}} &= \gamma_2 \overline{x}_{(t-1)n_h} + (1-\gamma_2) E_{(t-2)x_{n_h}} \\ B''_{(t-1)n_h} &= \gamma_1 E_{(t-1)x_{n_h}} + (1-\gamma_1) B''_{(t-2)n_h} \end{aligned} \tag{8}$$

where $\overline{x}_{(t-1)n_h}$ indicates sample mean of the variable x, which is based on the sample S_{n_h} at time t-1.

d. HEWMA statistics for the variable y, which is based on the sample S_{u_h} of size u_h (h = 1, 2, …., K) has been defined as

$$\begin{aligned} E_{ty_{u_h}} &= \gamma_2 \overline{y}_{tu_h} + (1-\gamma_2) E_{(t-1)y_{u_h}} \\ A''_{tu_h} &= \gamma_1 E_{ty_{u_h}} + (1-\gamma_1) A''_{(t-1)u_h} \end{aligned} \tag{9}$$

where $\overline{y}_{tu_h}$ indicates sample mean of the variable y, which is based on the sample S_{u_h} at time t.

The statistics draw up in equations (6), (7), (8) and (9) are unbiased estimators for the population mean $\overline{Y}_h$, $\overline{X}_h$, $\overline{X}_h$ and $\overline{Y}_h$ respectively.

Using the above HEWMA statistics defined in equation (9), we have constructed a memory type estimator for population mean based on the sample of size u_h taken afresh on the second occasion, which is presented below:

$$T'_{tu_h} = A''_{tu_h} \left[\frac{\overline{Z}_h}{\overline{z}_{u_h}} \right] \tag{10}$$

Therefore, the estimator of the overall population mean, i.e., $\overline{Y}$, which is based on the sample S_{u_h} of size u_h at time t, is described as

$$T_{tu} = \sum_{h=1}^{K} W_h T'_{tu_h} \tag{11}$$

where W_h is the stratum's weight.

Motivated by Kiregyra (1984) and using the above HEWMA statistics defined in equations (6), (7) and (8), we have constructed another memory type estimator for population mean based on the sample size m_h common for both the occasions, which is defined as

$$T''_{tm_h} = A'_{tm_h} + \beta_{yx_h}\left[\left(B''_{tn_h} - B'_{tm_h}\right) + \beta_{xz_h}\left(\bar{z}_{n_h} - \bar{Z}_h\right)\right] \tag{12}$$

Therefore, the estimator of the overall population mean, i.e., $\bar{Y}$, which is based on the sample S_{m_h} of size m_h at time t is described as

$$T_{tm} = \sum_{h=1}^{K} W_h T''_{tm_h} \tag{13}$$

where W_h is the stratum's weight.

Combining the estimator T_{tu} and T_{tm}, we get the final estimator of the population mean as follows:

$$T_{pe} = \emptyset T_{tu} + (1-\emptyset)T_{tm} \tag{14}$$

5. Variance of the proposed estimator T_{pe}

The variance of the proposed estimator T_{pe} is defined as

$$V(T_{pe}) = E\left(T_{pe} - \bar{Y}\right)^2 = E\left[\emptyset\left(T_{tu} - \bar{Y}\right) + (1-\emptyset)\left(T_{tm} - \bar{Y}\right)\right]^2$$

$$= \emptyset^2 V(T_{tu}) + (1-\emptyset)^2 V(T_{tm}) + 2\emptyset(1-\emptyset)Cov(T_{tu}, T_{tm})$$

where $V(T_{tu}) = E\left(T_{tu} - \bar{Y}\right)^2, V(T_{tm}) = E\left(T_{tm} - \bar{Y}\right)^2$

and $Cov(T_{tu}, T_{tm}) = E\left(T_{tu} - \bar{Y}\right)\left(T_{tm} - \bar{Y}\right)$

where, T_{tu} and T_{tm} are based on two independent samples u and m, so their covariance term is zero., i.e., $Cov(T_{tu}, T_{tm}) = 0$

Therefore, $V\left(T_{pe}\right) = \emptyset^2 V(T_{tu}) + (1-\emptyset)^2 V(T_{tm})$ (15)

Variance of the estimator T_{tu} is defined as

$$\begin{aligned} V(T_{tu}) &= E\left(T_{tu} - \bar{Y}\right)^2 \\ &= \sum_{h=1}^{K} W_h^2 V\left(T'_{tu_h}\right) \\ &= \sum_{h=1}^{K} W_h^2 E\left(T'_{tu_h} - \bar{Y}_h\right)^2 \end{aligned} \tag{16}$$

Similarly, the variance of the estimator T_{tm} is defined as

$$V(T_{tm}) = E\left(T_{tm} - \overline{Y}\right)^2$$
$$= \sum_{h=1}^{K} W_h^2 V\left(T''_{tm_h}\right)$$
$$= \sum_{h=1}^{K} W_h^2 E\left(T''_{tm_h} - \overline{Y}_h\right)^2 \tag{17}$$

To obtain the variance, we use the following transformation:

$$A'_{tm_h} = \overline{Y}_h(1+e_0), B'_{tm_h} = \overline{X}_h(1+e_1), A''_{tu_h} = \overline{Y}_h(1+e_2),$$
$$\overline{z}_{u_h} = \overline{Z}_h(1+e_3), B''_{tn_h} = \overline{X}_h(1+e_4), \overline{z}_{n_h} = \overline{Z}_h(1+e_5)$$

Such that

$$E(e_i) = 0 \text{ and } |e_i| < 1 \; \forall \; i = 0,1,2,3,4,5.$$

The expected values of the parameters under the above transformation are obtained as

$$E(e_0^2) = f_{1h} C_{y_h}^2 \frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2} R, \qquad E(e_1^2) = f_{1h} C_{x_h}^2 \frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2} R, \qquad E(e_2^2) = f_{2h} C_{y_h}^2 \frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2} R,$$

$$E(e_3^2) = f_{2h} C_{z_h}^2 \frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2} R, \qquad E(e_4^2) = f_{3h} C_{x_h}^2 \frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2} R, \qquad E(e_5^2) = f_{3h} C_{z_h}^2 \frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2} R,$$

$$E(e_0e_1) = f_{1h} \rho_{yx_h} C_{y_h} C_{x_h} \frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2} R, \qquad E(e_0e_4) = f_{3h} \rho_{yx_h} C_{y_h} C_{x_h} \frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2} R, \tag{18}$$

$$E(e_0e_5) = f_{3h} \rho_{yz_h} C_{y_h} C_{z_h} \frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2} R, \qquad E(e_1e_4) = E(e_4^2) = f_{3h} C_{x_h}^2 \frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2} R,$$

$$E(e_1e_5) = E(e_4e_5) = f_{3h} \rho_{xz_h} C_{x_h} C_{z_h} \frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2} R,$$

$$E(e_2e_3) = f_{2h} \rho_{yz_h} C_{y_h} C_{z_h} \frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2} R,$$

where,

$$f_{1h} = \frac{1}{m_h} - \frac{1}{N_h}, \quad f_{2h} = \frac{1}{u_h} - \frac{1}{N_h}, \quad f_{3h} = \frac{1}{n_h} - \frac{1}{N_h}$$

Expressing equation (10) in terms of e's we have

$$T'_{tu_h} = \overline{Y}_h(1+e_2)\left[\frac{\overline{Z}_h}{\overline{Z}_h(1+e_3)}\right]$$

Neglecting terms of e's having power greater than two, we have

$$T'_{tu_h} = \overline{Y}_h(1+e_2)\,(1-e_3+{e_3}^2)$$

It is found that

$$\left(T'_{tu_h} - \overline{Y}_h\right) \cong \left[\left\{\overline{Y}_h(1+e_2)\,(1-e_3+{e_3}^2)\right\} - \overline{Y}_h\right] \tag{19}$$

Squaring both sides of equation (19) and neglecting terms of e's having power greater than two and taking expectation using the result form equation (18), we have

$$\mathrm{E}\left(\mathrm{T}'_{\mathrm{tu_h}} - \overline{\mathrm{Y}}_{\mathrm{h}}\right)^2 = \left\{\mathrm{C}^2_{\mathrm{z_h}} - 2\rho_{\mathrm{yz_h}}\mathrm{C}_{\mathrm{y_h}}\mathrm{C}_{\mathrm{z_h}} + \mathrm{C}^2_{\mathrm{y_h}}\right\}{\overline{\mathrm{Y}}_{\mathrm{h}}}^2 \mathrm{f}_{2\mathrm{h}}\frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2}\mathrm{R} \tag{20}$$

Substituting the expression of $\mathrm{E}\left(\mathrm{T}'_{\mathrm{tu_h}} - \overline{\mathrm{Y}}_{\mathrm{h}}\right)^2$ into the equation (16), we get

$$\mathrm{V}(\mathrm{T_{tu}}) = \sum_{\mathrm{h}=1}^{\mathrm{K}} \mathrm{W}^2_{\mathrm{h}}\left\{\mathrm{C}^2_{\mathrm{z_h}} - 2\rho_{\mathrm{yz_h}}\mathrm{C}_{\mathrm{y_h}}\mathrm{C}_{\mathrm{z_h}} + \mathrm{C}^2_{\mathrm{y_h}}\right\}{\overline{\mathrm{Y}}_{\mathrm{h}}}^2 \mathrm{f}_{2\mathrm{h}}\frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2}\mathrm{R} \tag{21}$$

Similarly, expressing equation (12) in terms of e's, we have

$$\begin{aligned} T''_{tm_h} &= \overline{Y}_h(1+e_0) + \beta_{yx_h}\left[\overline{X}_h(1+e_4) - \overline{X}_h(1+e_1) + \beta_{xz_h}\left\{\overline{Z}_h(1+e_5) - \overline{Z}_h\right\}\right] \\ &= \overline{Y}_h(1+e_0) + \beta_{yx_h}\left[\overline{X}_h(e_4-e_1) + \beta_{xz_h}\overline{Z}_h e_5\right] \end{aligned}$$

It is found that

$$T''_{tm_h} - \overline{Y}_h = \left\{\overline{Y}_h(1+e_0) + \beta_{yx_h}\left[\overline{X}_h(e_4-e_1) + \beta_{xz_h}\overline{Z}_h e_5\right]\right\} - \overline{Y}_h \tag{22}$$

Squaring both sides of equation (22) and neglecting terms of e's having power greater than two and taking expectation using the result form equation (18), we have

$$E\left(T''_{tm_h} - \overline{Y}_h\right)^2 = \left\{(1-\rho^2_{yx_h})f_{1h} + \left(\rho_{yx_h} + 2\rho_{xz_h}\rho_{yz_h} + \rho_{yx_h}\rho^2_{xz_h}\right)\rho_{yx_h}f_{3h}\right\}S^2_{y_h}\frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2}\mathrm{R} \tag{23}$$

Substituting the expression of $\mathrm{E}\left(\mathrm{T}''_{\mathrm{tm_h}} - \overline{\mathrm{Y}}_{\mathrm{h}}\right)^2$ into the equation (17), we get

$$V(T_{tm}) = \sum_{h=1}^{K}\left[\begin{array}{c} W_h^2\left\{(1-\rho^2_{yx_h})f_{1h} + \left(\rho_{yx_h} + 2\rho_{xz_h}\rho_{yz_h} + \rho_{yx_h}\rho^2_{xz_h}\right)\rho_{yx_h}f_{3h}\right\} \\ S^2_{y_h}\frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2}R \end{array}\right] \tag{24}$$

Substituting the expression of $\mathrm{V}(\mathrm{T_{tu}})$ and $\mathrm{V}(\mathrm{T_{tm}})$ in the equation (15), we get

$$\begin{aligned} V(T_{pe}) = \emptyset^2 &\left[\sum_{h=1}^{K} W_h^2\left\{C^2_{z_h} - 2\rho_{yz_h}C_{y_h}C_{z_h} + C^2_{y_h}\right\}{\overline{Y}_h}^2 f_{2h}\frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2}R\right] \\ &+ (1-\emptyset)^2\sum_{h=1}^{K}\left[\begin{array}{c} W_h^2\left\{(1-\rho^2_{yx_h})f_{1h} + \left(\rho_{yx_h} + 2\rho_{xz_h}\rho_{yz_h} + \rho_{yx_h}\rho^2_{xz_h}\right)\rho_{yx_h}f_{3h}\right\} \\ S^2_{y_h}\frac{(\gamma_2\gamma_1)^2}{(\gamma_2-\gamma_1)^2}R \end{array}\right] \end{aligned} \tag{25}$$

6. Minimum variance the proposed estimator T_{pe}

We see that the variance of the estimator T_{pe} derived in equation (25) is a function of unknown constant $\emptyset$. So, to get the minimum (optimum) value of $\emptyset$, we differentiated equation (25) with respect to $\emptyset$ and equated the outcome to zero, which gives us the minimum (optimum) value of $\emptyset$ as

$$\emptyset_{opt} = \frac{V(T_{tm})}{V(T_{tu})+V(T_{tm})} \tag{26}$$

Substituting the optimum value of $\emptyset$ in the equation (15), we obtain the minimum (optimum) variance of the estimator T_{pe} as

$$V\left(T_{pe}\right)_{opt} = \frac{V(T_{tu}).V(T_{tm})}{V(T_{tu})+V(T_{tm})} \tag{27}$$

Substitute the expression of $V(T_{tu})$ and $V(T_{tm})$ in the equation (27), we get the expression of minimum variance of the estimator T_{pe}.

7. Empirical study

To test the performance of the proposed estimator, we have compared our estimator with the conventional sample mean estimator of population mean $\overline{Y}$ based on sample of size n_h $(h = 1,2, \ldots, K)$given by

$$\text{i. e.}, T_1 = \sum_{h=1}^{K} W_h^2\, \overline{y}_{n_h}$$

and its variance is obtained as

$$Var(T_1) = \textstyle\sum_{h=1}^{K} \left(\frac{1}{n_h} - \frac{1}{N_h}\right) W_h^2\, \overline{Y}_h^{\,2}\, C_{y_h}^2 \tag{28}$$

We have examined the merits of the proposed estimator against the conventional one through artificially generated population data as well as natural population data set. For this purpose, we have calculated the PRE of the proposed estimator with respect to T_1 as

$$PRE = \frac{Var(T_1)}{V\left(T_{pe}\right)_{opt}} \times 100 \tag{29}$$

7.1. Numerical illustration using artificially generated population

The main perspective of simulation study is that it replicates the actual system. Simulation helps us to compare the efficiency through artificial population generation technique and conclude whether a newly developed method is superior than the existing ones. Inspired by the work of Singh and Deo (2003), Sing *et al.* (20017), Maji *et al.* (2019) of artificially population generation techniques, we have generated three sets of independent random numbers (x, y, z) of size N (N=100) i.e., x[k], y[k] and z[k] (k = 1, 2, 3, …, N) from a standard normal distribution by using statistical software R.

For generating the population artificially, we use the following algorithm:

1. Generate random variables x1, y1, z1 and a (temporary variables) which are normally distributed with mean 0, S.D. =1 and are of size 100.
2. Define

   N =100.
3. Define

   x = 0, y = 0, z = 0, ry1x1 = 0.5 (correlation coefficient between y1 and x1), rx1z1=0.75 (correlation coefficient between x1 and z1), Sx1 = $\sqrt{50}$,

   Sy1 = $\sqrt{50}$, Sa = $\sqrt{40}$

   (Sx1, Sy1, Sa are S. D. (standard deviations) of x1, y1, z1 and variable a respectively)

   mx1= 20 (mean of x1), mz1 = 25 (mean of z1).
4. Define

   a1=Sy1*Sy*(1-(ry1x1^2))

   a2=Sz*Sz*(1-(rx1z1^2))
5. for (j in 1 to N)

   {

   y[j]=20.0+(sqrt(a1)*y1[j])+(ry1x1*sy1*x1[j])

   x[j]=25.0+(sx1*x1[j])

   z[j]=15+(sqrt(a2)*z1[j])+(rx1z1*sz1*x1[j])

   }
6. Take output of the variables x, y and z.

In such a way artificial population data set 1 has been generated.

Repeating the above algorithm, artificial population data set 2 has been generated with the following changed ry1x1 = 0.75 and rx1z1=0.5 in step 3.

Similarly, artificial population data set 3 has been generated by changing the value of ry1x1 = 0.75 and rx1z1=0.75 in step 3.

By the above algorithm we have generated artificial population of size 100 for each population data set and each data set further divided into five strata sequentially with sizes 15, 18, 21, 22 and 24 respectively. Details are given below in Table 1.

We have calculated the PREs of the proposed estimator T_{pe} for the different values of γ_1 and γ_2 from the Artificial Data Set-I, Artificial Data Set-II, Artificial Data Set-III.

For generating the table, we have taken ρ_{yx} = ry1x1 and ρ_{xz}= rx1z1.

Table 1: Parameters of the stratum 1-5

Size	Simulated Data				
	Stratum 1	**Stratum 2**	**Stratum 3**	**Stratum 4**	**Stratum 5**
N_h	15	18	21	22	24
n_h	8	7	9	10	10
m_h	5	4	5	6	6
u_h	3	3	4	4	4

Table 2: PRE of the proposed estimator with respect to T_1 for simulated data sets

Constant		PRE		
γ_1	γ_2	**Artificial Data Set-I** ($\rho_{yx} = 0.50, \rho_{xz} = 0.75$)	**Artificial Data Set-II** ($\rho_{yx} = 0.75, \rho_{xz} = 0.50$)	**Artificial Data Set-III** ($\rho_{yx} = 0.75, \rho_{xz} = 0.75$)
0.06	0.05	1136.458	1003.131	1464.382
	0.25	624.383	551.132	804.549
	0.50	557.616	492.198	718.516
	0.75	532.230	469.789	685.804
	1.00	515.532	455.051	664.288
0.10	0.05	923.523	815.176	1190.004
	0.25	411.458	363.186	530.184
	0.50	344.727	304.284	444.197
	0.75	319.428	281.953	411.599
	1.00	302.942	267.401	390.355
0.14	0.05	832.053	734.438	1072.141
	0.25	320.000	282.458	412.336
	0.50	253.309	223.591	326.401
	0.75	228.104	201.343	293.923
	1.00	211.831	186.980	272.955

7.2. Numerical Illustration using Natural Population data set

We have considered two natural population data sets to examine the merits of the proposed estimator T_{pe}. The sources of populations and details of the variables y, x and z and the values of various parameters are mentioned bellow.

7.2.1. Population Data Set-I [Literacy Rate by Sex of India, Census India (2011)]

y: Number of literates (male and female) in the year 2011;
x: Number of literates (male and female) in the year 2001;
z: Female literacy rate (2011).

We have divided 32 states of India into 6 different strata (zone wise) as shown in Table 3.

Table 3: Values of different parameters of the respective variables

Strata	States	Statistical Parameters
Stratum I	Andhra Pradesh Karnataka Kerala Tamil Nadu	$N_h = 4, n_h = 3, m_h = 2, \bar{X}_h = 79.13, \bar{Y}_h = 72.86, \bar{Z}_h = 73.18, S^2_{y_h} = 127.7, S^2_{x_h} = 172.18, S^2_{z_h} = 194.02, \rho_{yx_h} = 0.9929, \rho_{yz_h} = 0.9980, \rho_{xz_h} = 0.9970$
Stratum II	Goa Gujarat Maharashtra Punjab Rajasthan	$N_h = 5, n_h = 4, m_h = 2, \bar{X}_h = 78.18, \bar{Y}_h = 71.62, \bar{Z}_h = 70.62, S^2_{y_h} = 69.82, S^2_{x_h} = 67.83, S^2_{z_h} = 142.49, \rho_{yx_h} = 0.9864, \rho_{yz_h} = 0.9864, \rho_{xz_h} = 0.9759$
Stratum III	West Bengal Odisha Jharkhand Bihar	$N_h = 4, n_h = 3, m_h = 2, \bar{X}_h = 69.95, \bar{Y}_h = 58.07, \bar{Z}_h = 60.35, S^2_{y_h} = 40.20, S^2_{x_h} = 93.24, S^2_{z_h} = 73.06, \rho_{yx_h} = 0.9997, \rho_{yz_h} = 0.9900, \rho_{xz_h} = 0.9930$
Stratum IV	Manipur Meghalaya Nagaland Arunachal Pradesh Assam Sikkim Tripura	$N_h = 7, n_h = 4, m_h = 2, \bar{X}_h = 76.73, \bar{Y}_h = 65.61, \bar{Z}_h = 71.66, S^2_{y_h} = 49, S^2_{x_h} = 34.95, S^2_{z_h} = 64.02, \rho_{yx_h} = 0.9266, \rho_{yz_h} = 0.9702, \rho_{xz_h} = 0.8663$
Stratum V	Uttar Pradesh Haryana Himachal Pradesh Uttarakhand Jammu & Kashmir	$N_h = 5, n_h = 4, m_h = 2, \bar{X}_h = 74.42, \bar{Y}_h = 65.56, \bar{Z}_h = 65.08, S^2_{y_h} = 47.02, S^2_{x_h} = 87.15, S^2_{z_h} = 69.85, \rho_{yx_h} = 0.9987, \rho_{yz_h} = 0.9981, \rho_{xz_h} = 0.9941$
Stratum VI	A & N Island Chandigarh Daman & Diu D & N Haveli Delhi Lakshadweep Pondicherry	$N_h = 7, n_h = 4, m_h = 2, \bar{X}_h = 85.67, \bar{Y}_h = 78.37, \bar{Z}_h = 79.54, S^2_{y_h} = 21.76, S^2_{x_h} = 89.89, S^2_{z_h} = 52.67, \rho_{yx_h} = 0.9532, \rho_{yz_h} = 0.9822, \rho_{xz_h} = 0.9862$

7.2.2. Population Data Set-II [Abortion Rate, Statistical Abstract of the United States (2011)]

y: Number of abortions reported in the year 2008.
x: Number of abortions reported in the year 2007.
z: Number of abortions reported in the year 2005.

We have divided 51 states of the United States into 4 different strata (zone wise) as shown in Table 4.

Table 4: Values of different parameters of the respective variables

Strata	States	Statistical Parameters
Stratum I	Wyoming	$N_h = 14, n_h = 8, m_h = 5, u_h = 3, \overline{X}_h = 6.551$, $\overline{Y}_h = 6.59$, $\overline{Z}_h = 6.720$, $S^2_{y_h} = 4.56$, $S^2_{x_h} = 4.51$, $S^2_{z_h} = 5.21$, $\rho_{yx_h} = 0.9784$, $\rho_{yz_h} = 0.9725$, $\rho_{xz_h} = 0.9484$
	Missouri	
	Mississippi	
	Kentucky	
	Oklahoma	
	Arkansas	
	Indiana	
	Nebraska	
	South Carolina	
	Wisconsin	
	Utah	
	South Dakota	
	Idaho	
	West Virginia	
Stratum II	Alaska	$N_h = 25$, $n_h = 12$, $m_h = 7$, $u_h = 5, \overline{X}_h = 15.031$, $\overline{Y}_h = 15.11$, $\overline{Z}_h = 14.851$, $S^2_{y_h} = 6.91$, $S^2_{x_h} = 6.93$, $S^2_{z_h} = 8.87$, $\rho_{yx_h} = 0.9413$, $\rho_{yz_h} = 0.878$, $\rho_{xz_h} = 0.8988$
	Montana	
	New Hampshire	
	Minnesota	
	Vermont	
	Ohio	
	Arizona	
	New Mexico	
	North Dakota	
	Maine	
	Michigan	
	Massachusetts	
	Washington	
	Kansas	
	Virginia	
	North Carolina	
	Oregon	
	Pennsylvania	
	Texas	
	Louisiana	
	Colorado	
	Tennessee	
	Iowa	
	Alabama	
	Georgia	

Table 4: Values of different parameters of the respective variables (cont.)

Strata	States	Statistical Parameters
Stratum III	Hawaii	$N_h = 7, n_h = 4, m_h = 2, u_h = 2, \overline{X}_h = 24.562, \overline{Y}_h = 24.48, \overline{Z}_h = 24.095, S^2_{y_h} = 31.42, S^2_{x_h} = 37.22, S^2_{z_h} = 65.29, \rho_{yx_h} = 0.9885, \rho_{yz_h} = 0.9442, \rho_{xz_h} = 0.95454$
	Rhode Island	
	Connecticut	
	Nevada	
	Florida	
	California	
	Illinois	
Stratum IV	Maryland	$N_h = 5, n_h = 4, m_h = 2, u_h = 2, \overline{X}_h = 33.528, \overline{Y}_h = 33.55, \overline{Z}_h = 36.533, S^2_{y_h} = 19.31, S^2_{x_h} = 29.03, S^2_{z_h} = 53.75, \rho_{yx_h} = 0.9751, \rho_{yz_h} = 0.39256, \rho_{xz_h} = 0.53965$
	District of Columbia	
	New Jersey	
	New York	
	Delaware	

We have calculated the PREs of the proposed estimator T_{pe} for the different values of γ_1 and γ_2 from the Population Data Set-I and Population Data Set-II.

Table 5. PRE of the proposed estimator with respect to T_1 for Natural Population data sets

Constant		PRE	
γ_1	γ_2	Population Data Set-I	Population Data Set-II
0.06	0.05	11392.639	4077.631
	0.25	6259.249	2240.298
	0.50	5589.930	2000.737
	0.75	5335.434	1909.648
	1.00	5168.048	1849.738
0.10	0.05	9258.025	3313.614
	0.25	4124.737	1476.318
	0.50	3455.776	1236.885
	0.75	3202.167	1146.114
	1.00	3036.894	1086.959
0.14	0.05	8341.070	2985.419
	0.25	3207.905	1148.167
	0.50	2539.341	908.876
	0.75	2286.668	818.440
	1.00	2123.543	760.054

8. Conclusion

We have examined the performances of our proposed strategy against the conventional ones for different data sets as presented. The results of such a comparison are discussed in Table 2 and Table 5. Thus, the following interpretations may be read out from the respective Tables.

a) It is noted that the development of the estimation procedure in successive sampling is still in progress under various designs but no one has incorporated the past sample information in the form of HEWMA statistics to construct an effective estimation strategy in the successive sampling scheme. The proposed estimator uses the current sample information along with the information available from past samples in the form of HEWMA statistics. The results in Tables 2 and 5 demonstrated that the proposed estimator is more efficient than the conventional estimator. It may be concluded that the suggested strategy is more efficient in estimation of population mean.

b) From Table 2, it is observed that for fixed values of ρ_{yx}, the values of PREs of our proposed estimator are increasing with the increasing values of ρ_{xz}. A similar pattern may be noted for fixed values of ρ_{xz} with increasing values of ρ_{yx}. Thus, it is clear that for the presence of higher values of correlation coefficient between study and auxiliary variable, our proposed strategy produces more precise estimates.

   From Table 5, it is also noted that the correlation coefficient between variables (y and x) and (x and z) is high in most of the strata and consequently the PREs are very high. Therefore, the findings of the simulation studies are also justified with the natural population studies.

c) It may also be noted that in artificially generated population, values of the various statistical parameter such as means, variances, correlation coefficients etc. are almost strata wise similar while in case of natural populations, their parametric values are different from strata to strata. Our proposed estimator performs profoundly for both the types of population which enhance their recommendation in practice.

Thus, it is observed from the interpretation of the result that the use of memory-based information of HEWMA statistics for the estimation of population mean under stratified successive sampling is highly encouraging. Moreover, the calibration technique utilized in the formulation of the estimator helps us in enhancing the efficiency of the proposed strategy. Therefore, looking at the pleasing findings, we are recommended our proposed strategy to the survey practitioners for their application in real life.

Acknowledgement

Authors are thankful to the reviewers for their constructive suggestions, which enhanced the quality of the manuscript.

References

Aslam, I., Noor-ul-Amin, M., Yasmeen, U. and Hanif, M., (2020). Memory type ratio and product estimators in stratified sampling. *Journal of Reliability and Statistical Studies*, 13(1), pp. 1–20.

Aslam, I., Noor-ul-Amin, M., Hanif, M. and Sharma, P., (2023). Memory type ratio and product estimators under ranked-based sampling schemes. *Communications in Statistics-Theory and Methods*, 1–23. *Communications in Statistics - Theory and Methods,* 52, pp. 1–23.

Bhushan, S.; Kumar, A., Al-Omari, A. I. and Alomani, G. A., (2023). Mean Estimation for Time-Based Surveys Using Memory-Type Logarithmic Estimators. *Mathematics*, 1(9), pp. 21–25. https://doi.org/10.3390/math11092125.

Biradar, R. S., Singh, H. P., (2001). Successive sampling using auxiliary information on both occasions. *Calcutta Statistical Association Bulletin,* 51, pp. 243–251.

Chaturvedi, D. K., Tripathi, T. P., (1983). Estimation of population ratio on two occasions using multivariate auxiliary information. *Journal of Indian Statistical Association*, 21, pp. 113–120.

Das, A. K., (1982). Estimation of population ratio on two occasions. *Journal of the Indian Society of Agricultural Statistics*, 34, pp. 1–9.

Feng, S., Zou, G., (1997). Sample rotation method with auxiliary variable. *Communications in Statistics-Theory and Methods*, 26, 6, pp. 1497–1509.

Haq, A., (2013). A new hybrid exponentially weighted moving average control chart for monitoring process mean. *Quality and Reliability Engineering International,* 29(7), pp. 1015–1025.

Haq, A., (2016). A new hybrid exponentially weighted moving average control chart for monitoring process mean Discussion. *Quality and Reliability Engineering International*, 33(7), pp. 629–1631.

Jessen, R. J., (1942). Statistical Investigation of a Sample Survey for obtaining farm facts. *Iowa Agricultural Experiment Station Research Bulletin*, No. 304, Ames, Iowa, USA, pp. 1–104.

Kiregyera, B., (1984). Regression-type estimators using two auxiliary variables and the model of double sampling from finite populations. *Metrika,* 31, pp. 215–226.

Maji, R., Singh, G. N. and Bandyopadhyay, A., (2019). Estimation of Population Mean in Presence of Random Non-Response in Two-Stage Cluster Sampling. *Communications in Statistics - Theory and Methods,* 48 (14), pp. 3586–3608.

Noor-ul-Amin, M., (2020) Memory type ratio and product estimators for population mean for time-based surveys. *Journal of Statistical Computation and Simulation*, 90(17), pp. 3080–3092.

Noor-ul-Amin, M., (2021). Memory type estimators of population mean using exponentially weighted moving averages for time scaled surveys. *Communications in Statistics-Theory and Methods*, 50(12), pp. 2747–2758.

Patterson, H. D., (1950). Sampling on successive occasions with partial replacement of units. *Journal of the Royal Statistical Society*, 12, pp. 241–255.

Rao, J. N. K., Graham, J. E., (1964). Rotation design for sampling on repeated occasions. *Journal of the American Statistical Association*, 59, pp. 492–509.

Roberts, S., (1959). Control chart tests based on geometric moving averages. *Techno metrics*, 1(3), pp. 239-250.

Sen, A. R., (1971). Successive sampling with two auxiliary variables. *Sankhya*, 33, Series B, pp. 371–378.

Sen, A. R., (1972). Successive sampling with p (p ≥1) auxiliary variables. *Annals Mathematical Statistics*, 43, pp. 2031–2034.

Sen, A. R., (1973). Theory and application of sampling on repeated occasions with several auxiliary variables. *Biometrics*, 29, pp. 381–385.

Singh, G. N., Singh, V. K., (2001). On the use of auxiliary information in successive sampling. *Journal of the Indian Society of Agricultural Statistics*, 54(1), pp. 1–12.

Singh, G. N., (2003). Estimation of population mean using auxiliary information on recent occasion in h-occasion successive sampling. *Statistics in Transition*, pp. 523-532.

Singh, S., Deo, B., (2003). Imputation by power transformation. *Statistical Papers*, 4, pp. 555–579.

Singh, G. N., Karna, J. P., (2009). Estimation of population mean on current occasion in two-occasion successive sampling. *Metron - International Journal of Statistics*, vol. LXVII, no. 1, pp. 87–103.

Singh, H. P., Vishwakarma, G. K., (2009). A general procedure for estimating population mean in successive sampling. *Communications in Statistics-Theory and Methods,* 38(2), pp. 293–308.

Singh, H. P., Tailor, R., Singh, S. and Kim, J. M., (2011). Estimation of population variance in successive sampling. *Quality and Quantity*, 45, pp. 477–494.

Singh, G. N., Sharma, A. K. and Bandyopadhyay, A., (2017). Effectual Variance Estimation Strategy in Two Occasions Successive Sampling in Presence of Random Non-Response. *Communications in Statistics-Theory and Methods*, 46(14), pp. 7201–7224.