**Academic Journal of Statistic and mathematics**
Vol.12, No.4; April-2026;
ISSN (5730–7151);
p –ISSN 4052–392X
Impact factor: 6.53

# OPTIMAL WEIGHT ADJUSTMENT WITH SUBSAMPLING THE NONRESPONDENTS IN LONGITUDINAL SURVEY FOR SMALL AREA ESTIMATION

**Iseh, Matthew Joshua Bassey, Mbuotidem Okon and *Etebong P. Clemen***
Department of Statistics, Akwa Ibom State University, Mkpat Enin, Nigeria
**ORCID ID: https://orcid.org/0000-0003-2696-7319**
Department of Statistics, Akwa Ibom State University, Mkpat Enin, Nigeria
Department of Statistics, University of Uyo, Nigeria,

**Abstract:** Efficient estimation of population parameters in small area estimation (SAE) is crucial, especially under stratified random sampling when nonresponse occurs. Several estimators have been developed to improve estimation accuracy. However, some of these methods often fail to fully adjust for nonresponse bias, leading to unreliable domain mean estimates. This study proposes a new calibration-based estimator, by integrating the Kullback-Leibler distance function-based weight adjustments. The estimator is formulated under two conditions: when nonresponse affects only the study variable, and when nonresponse affects both study and auxiliary variables. Properties of the estimators have been derived and the results confirm that the proposed estimator provides greater efficiency and lower error rates than existing methods. Empirical validation is conducted using data from a longitudinal survey (before, during, and after COVID-19) from Household Finance & Consumption Survey (HFCS) and Integrated Household Survey (IHS) for 2019-2021 from the Department of Statistic of the Central Bank of Nigeria. Comparative performance analysis using variance and mean square error calculations demonstrates that the proposed estimator consistently outperforms the existing domain estimators in handling nonresponse across different domains. The study concludes that calibration techniques enhance estimation accuracy in stratified sampling and offers recommendations for further research on alternative calibration functions in small area estimation.

**Keywords**: Calibration estimator, Kullback-Leibler Distance measure, Mean square error, Nonresponse, Small area estimation, Stratified random sampling

## 1. **Introduction**

Accurate estimation of population parameters for small areas or domains is essential in statistical surveys, particularly when using stratified random sampling. In many applications, such as economic planning, health studies, and social research, data must be disaggregated to provide reliable estimates for sub-populations or geographic regions. However, small sample sizes within strata often lead to unreliable direct estimators with high variances. To address this challenge, Small Area Estimation (SAE) techniques have been developed to improve estimation accuracy by leveraging auxiliary information and model-based approaches (Rao, 2003; Datta, 2009). Among these, calibration estimators have gained prominence due to their ability to reduce nonresponse bias and improve efficiency in stratified sampling (Iseh & Bassey, 2022; Iseh & Bassey, 2024; Iseh et al., 2025).

Several studies have explored domain estimation under stratified sampling, focusing on direct and indirect estimation methods. Direct estimators rely solely on within-domain data, making them impractical when sample sizes are insufficient (Jiang & Lahiri, 2006). Indirect estimators, such as synthetic estimators, borrow

strength from larger domains but introduce biases due to incorrect model assumptions (Gonzalez, 1973; Sarndal, 1984; Iseh & Enang, 2021). More recent studies have introduced calibration estimators that use auxiliary information to minimize variance and bias in stratified samples (Khare et al., 2021; Iseh & Bassey, 2021a and Iseh & Bassey, 2021b). Despite these advancements, existing methods still struggle with nonresponse and the optimal adjustment of design weights for unbiased domain mean estimation (Iseh & Bassey, 2024 and Ikot & Iseh, 2024).
The primary challenge in small area estimation under stratified sampling arises when survey nonresponse leads to missing data within certain strata, reducing the reliability of direct estimators. Traditional SAE models either fail to adjust for nonresponse or rely on assumptions that may not hold for real-world data (Singh & Sisodia, 2011). Many existing approaches assume that auxiliary variables are fully observed, which is often unrealistic in longitudinal surveys (Ashutosh, 2021; Bassey & Iseh, 2025). There is a need for a refined calibration-based estimator that effectively handles both small sample sizes and nonresponse, ensuring improved precision while maintaining minimal bias.
This study introduces a new calibration estimator for stratified sampling by extending the Hansen and Hurwitz (1946) framework to adjust design weights using a Kullback-Leibler distance function. The proposed estimators compensate for nonresponse by aligning design weights with calibrated weights while incorporating supplementary information from auxiliary variables. The emphasis of this study is to evaluate both analytically and empirically how efficient the proposed estimators are, relative to the unbiased Hansen and Hurwitz (1946) estimator and other existing estimators (Tikkiwal & Ghiya (2000a), Tikkiwal & Ghiya (2000b), Ashutosh (2021) and Ikot and Iseh, (2024)) in domain estimation. Empirical validation is conducted using household expenditure data from the Central Bank of Nigeria, demonstrating the estimator's applicability in real-world stratified survey contexts.

## 2. Sampling Strategy

Consider a stratified random sampling design with $H$strata and such that $n_h$ elements are considered from $N_h$ in stratum $h$, $h = 1, 2, 3, \ldots, H$. Then, the design weights needed for the point estimation are $W_{dh} = \frac{N_{dh}}{N_d}$ for all d-domains in stratum $h = 1,2,3,\ldots,N_h$. However, the design weights $W_{dh}$ needed for the variance estimation if $h \neq g$ and both $h$ and $g$ are in stratum $h$ is: $W_{dh} = \frac{N_{dh}(N_{dh}-1)}{N_d(N_d-1)}$.
Also considering the finite population under study population $U$ of size $N$ divided into $D$ domains; $U_1, U_2, U_3, \ldots, U_D$ of sizes; $N_1, N_2, N_3, \ldots, N_D$ respectively. It is assumed that domains may be quite large or small. Again, Hansen and Hurwitz (1946) suggested drawing a subsample of size $r = \frac{n_2}{g}$, where $g \geq 1$ is predetermined, from the $n_2 = n - n_1$ nonrespondents and eliciting responses from all of them. It is assumed that domains may be quite large or small. For a typical $d^{th}$domain $U_d$ several characteristics may be defined including the domain total, mean, variance and covariance respectively as

$Y_{U_d} = \sum_{U_d} y_{dg}$ , $\bar{Y}_{U_d} = \frac{1}{N_d}\sum_{U_d} y_{dg}$, $S^2_{U_d}(Y) = \frac{1}{N_d-1}\sum_{g\in U_d}(y_{dg} - \bar{Y}_{U_d})^2$, and
$C_{U_d}(X,Y) = \frac{1}{N_d-1}\sum_{g\in U_d}(x_{dg} - \bar{X}_{U_d})(y_{dg} - \bar{Y}_{U_d})$.

### 2.1. Some Existing Estimators under Stratified Sampling

Some existing estimators for population mean in stratified sampling with sub-sampling the non-respondents are presented as follows:

### 2.1.1 Hansen and Hurwitz (1946) Estimator

Hansen and Hurwitz (1946) proposed an unbiased estimator defined as:

$$\hat{\hat{y}}_d = \sum_{h=1}^{H} W_{dh}\bar{y}_{dh}^{*} \qquad (1)$$

where $\bar{y}_{dh}^{*} = \frac{n_{d1}\bar{y}_{d1} + n_{d2}\bar{y}_{d2}}{n_d}$,

$$V(\hat{\hat{y}}_d) = \sum_{h=1}^{H}\left(\frac{1}{n_{dh}} - \frac{1}{N_d}\right) W_{dh}^2 S_{ydh1}^2 + \sum_{h=1}^{H}\frac{K_{dh}-1}{n_{dh2}} W_{dh}^2 W_{dh2} S_{ydh2}^2$$

$S_{ydh1}^2$ and $S_{ydh2}^2$ are the mean square of the response and non-response groups respectively of the study variable in the population for the $h^{th}$ stratum and

$W_{dh} = \frac{N_{dh}}{N_d}$ is the response rate of the d domain in the $h^{th}$ stratum.

$W_{dh2} = \frac{N_{dh2}}{N_d}$ is the non-response rate of the $d^{th}$ domain in the $h^{th}$ stratum

.

### 2.1.2 Tikkiwal and Ghiya(2000a) Estimator

Tikkiwal and Ghiya (2000a) proposed a direct ratio estimator with stratified sampling as:

$$T_{DR.st.d} = \frac{\bar{y}_{st.d}}{\bar{x}_{st.d}}\bar{X}_{h.d} \qquad (2)$$

where $d^{th}$ domain mean of y: $\bar{y}_{st.d} = \sum_{h=1}^{H} W_{h.d}\,\bar{y}_{h.d}$
$d^{th}$ domain mean of x: $\bar{x}_{st.d} = \sum_{h=1}^{H} W_{h.d}\,\bar{x}_{h.d}$
Bias and mean square error of $T_{DR.st.d}$ are respectively given as follows

$$\text{Bias}(T_{DR.st.d}) = \left[\sum_{h=1}^{H} W_{h.d}\,\bar{Y}_{h.d}\frac{N_{h.d}-n_{h.d}}{N_{h.d}n_{h.d}} C_{YXh.d} - \bar{Y}_d\right]$$

$$\text{MSE}(T_{DR.st.d}) = \sum_{h=1}^{H} W_{h.d}^2\bar{Y}_{h.d}^2\frac{N_{h.d}-n_{h.d}}{N_{h.d}n_{h.d}}\left(C_{Yh.d}^2 + C_{Xh.d}^2 - 2C_{YXh.d}\right)$$

### 2.1.3 Tikkiwal and Ghiya (2000b) Estimator

The duo proposed a direct generalized estimator with the application of auxiliary variable stated as follows:

$$T_{DG.st.d} = \bar{y}_{st.d}\left[\frac{\bar{x}_{st.d}}{\bar{X}_{st.d}}\right]^{\delta} \qquad (3)$$

Bias and Mean Square Error of $T_{DG.st.d}$ are given respectively as:

$$\text{Bias}(T_{DG.st.d}) = \sum_{h=1}^{H} W_{h.d}\,\bar{Y}_{h.d}\frac{N_{h.d}-n_{h.d}}{N_{h.d}n_{h.d}}\left(\frac{\delta(\delta-1)}{2}C_{Xh.d}^2 + \delta C_{YXh.d}\right) - \bar{Y}_d$$

$$\text{MSE}(T_{DG.st.d}) = \sum_{h=1}^{H} W_{h.d}^2\bar{Y}_{h.d}^2\frac{N_{h.d}-n_{h.d}}{N_{h.d}n_{h.d}}\left(C_{Yh.d}^2 + \delta^2 C_{Xh.d}^2 + 2\delta C_{YXh.d}\right)$$

### 2.1.4 Ashutosh (2021) Estimator

The author proposed a direct generalized estimator for domain mean through stratified sampling with non-response as:

$$T_{DG.st.\beta.d}^{*} = \bar{y}_{st.d}^{*}\left[\frac{\bar{x}_{st.d}^{*}}{\bar{X}_{st.d}^{*}}\right]^{\beta} \qquad (4)$$

where $\beta$ is a chosen constant of $d^{th}$ domain,
$\bar{y}_{st.d}^{*} = \sum_{h=1}^{H} W_{h.d}\,\bar{y}_{h.d}$ is the $d^{th}$ domain mean of y, and
$\bar{x}_{st.d}^{*} = \sum_{h=1}^{H} W_{h.d}\,\bar{x}_{h.d}$ is the $d^{th}$ domain mean of x:
Members of the proposed family of estimators $T_{DG.st.\beta.d}^{*}$ are;

$T_{DG.st.0.d}^{*} = \bar{y}_{st.d}^{*}$ if $\beta = 0$

$T_{DG.st.-1.d}^{*} = \frac{\bar{y}_{st.d}^{*}}{\bar{x}_{st.d}^{*}}\bar{X}_{h.d}$ if $\beta = -1$

$T_{DG.st.1.d}^{*} = \bar{y}_{st.d}^{*}\frac{\bar{x}_{st.d}^{*}}{\bar{X}_{st.d}^{*}}$ if $\beta = 1$

$T^*_{DG.st.2.d} = \bar{y}^*_{st.d}\left[\frac{\bar{x}^*_{st.d}}{\bar{X}^*_{h.d}}\right]^2$ if $\beta = 2$

Bias and Mean Square Error of $T^*_{DG.st.\beta.d}$ are given respectively as:

$\text{Bias}(T^*_{DG.st.\beta.d}) = \sum_{h=1}^{H} W_{h.d}\,\bar{Y}_{h.d}\left[\frac{N_{h.d}-n_{h.d}}{N_{h.d}n_{h.d}}C^2_{Xh.d} + \frac{(g_{h.d}-1)W_{2h.d}}{n_{h.d}}C^2_{2Yh.d}\right] - \bar{Y}_d$

$\text{MSE}(T^*_{DG.st.\beta.d}) = \sum_{h=1}^{H} W^2_{h.d}\bar{Y}^2_{h.d}\left[\frac{N_{h.d}-n_{h.d}}{N_{h.d}n_{h.d}}\langle C^2_{Yh.d} + C^2_{Xh.d} + 2C_{YXh.d}\rangle + \frac{(g_{h.d}-1)W_{2h.d}}{n_{h.d}}\langle C\,^2_{2Yh.d} - 2C_{2YXh.d}\rangle\right]$

### 2.1.5 Ikot and Iseh (2024) Estimator:

Ikot and Iseh (2024) proposed calibration estimators for combined regression and ratio estimators under non-response as:

$t^*_{cal} = \sum_{h=1}^{L} W^*_{dh}\,\bar{y}_{dhr}$ (5)

where $\bar{y}_{dhr} = r_{dh}\bar{X}_{dh},$ and $r_{dh} = \frac{\bar{y}^*_{dh}}{\bar{x}_{dh}}$

$\text{Bias}(t^*_{cal}) = \sum_{h=1}^{L} W_{dh}\,\bar{Y}_{dh}\left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)\frac{S^2_{xdh}}{X^2_{dh}}\right] - \sum_{h=1}^{L} W_{dh}\,\bar{Y}_{dh}\left[\frac{1}{\bar{X}_{dh}}\frac{1}{\bar{Y}_{dh}}\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)\rho_{xy}S_{xdh}S_{ydh}\right]$

$\text{MSE}(t^*_{cal})_{min} = \sum_{h=1}^{L} W^2_{dh}\left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S^2_{ydh} + \frac{(K_{dh}-1)}{n_{dh2}}W_{dh2}S^2_{ydh2}\right] -$
$2\sum_{h=1}^{L} W^2_{dh}\frac{\bar{Y}_{dh}}{\bar{X}_{dh}}\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)\rho_{xy}S_{xdh}S_{ydh} + \sum_{h=1}^{L} W^2_{dh}\frac{\bar{Y}_{dh}}{\bar{X}_{dh}}\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S^2_{xdh} -$
$\frac{\left[\sum_{h=1}^{L} W^2_{dh}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)\rho_{xy}S_{xdh}S_{ydh} - \sum_{h=1}^{L} W^2_{dh}\frac{\bar{Y}_{dh}}{\bar{X}_{dh}}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)S^2_{xdh}\right]^2}{\sum_{h=1}^{L} W^2_{dh}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)S^2_{xdh}} - \frac{\left[\sum_{h=1}^{L} W^2_{dh}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)S_{xdh}S_{ydh}\lambda_{12} - \sum_{h=1}^{L} W^2_{dh}\frac{\bar{Y}_{dh}}{\bar{X}_{dh}}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)S^2_{xdh}\lambda_{03}\right]^2}{\sum_{h=1}^{L} W^2_{dh}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)S^4_{dh}(\lambda_{04}-1)}$

## 2.0 Proposed Estimators

Motivated by Hansen and Hurwitz (1946) as shown in (1), the proposed longitudinal estimator is defined as

$\hat{\bar{y}}^{(t)}_{st} = \sum_{h=1}^{L} W^{(t)}_{dh}\,\bar{y}^{*(t)}_{dh}$ (6)

where $\bar{y}^{*(t)}_{dh} = \frac{n_{d1}\bar{y}^{(t)}_{1d} + n_{d2}\bar{y}^{(t)}_{2d}}{n_d}$ and

$W^{(t)}_{dh}$ are the calibrated weights

The proposed longitudinal estimator in Eq. 6 above is calibrated upon two conditions:

**Condition 1**: The study variable $y$ is influenced by non-response but the auxiliary variable $x$ is free from non-response.

**Condition 2:** The study and auxiliary variables are not free from nonresponse.

Here, the Kullback distance measure is used to minimize the distance between the stratum weights $W_{dh}$ and the calibration weights $W^{(t)}_{dh}$ subject to the constraints under condition 1 and 2.

**Case I**: By adopting condition 1, the Kullback Distance function (φ) is given as

$\varphi = \sum_{h=1}^{L} W^{(t)}_{dh}\log\left(\frac{W^{(t)}_{dh}}{W_{dh}q_{dh}}\right) - W^{(t)}_{dh} - W_{dh}$

subject to the following calibration constraints

$\sum_{h=1}^{L} W^{(t)}_{dh}\,\bar{x}^{(t)}_{dh} = \bar{X}^{(t)}_d$

$\sum_{h=1}^{L} W^{(t)}_{dh}\,s^{2(t)}_{xdh} = S^{2(t)}_{xdh}$

Given the Lagrange function $L$, the optimization problem is given as

$$L=\sum_{h=1}^{L}W_{dh}^{(t)}\log\left(\frac{W_{dh}^{(t)}}{W_{dh}q_{dh}}\right)-W_{dh}^{(t)}-W_{dh}-\gamma_1\left[\sum_{h=1}^{L}W_{dh}^{(t)}\bar{x}_{dh}^{(t)}-\bar{X}_d^{(t)}\right]$$
$$-\gamma_2\left[\sum_{h=1}^{L}W_{dh}^{(t)}s_{xdh}^{2(t)}-S_{xdh}^{2(t)}\right]$$

Taking partial derivative of $L$ with respect to $W_{dh}^{(t)}$, we have:

$$\Rightarrow W_{dh}^{(t)}=W_{dh}q_{dh}\left[1+\gamma_1\bar{x}_{dh}^{(t)}+\gamma_2 s_{xdh}^{2(t)}\right]$$

Hence Eq. 6 becomes

$$\hat{\bar{y}}_{st1}^{(t)}=\sum_{h=1}^{L}W_{dh}q_{dh}\,\bar{y}_{dh}^{*(t)}+\hat{\beta}_1\left(\bar{X}_d^{(t)}-\sum_{h=1}^{L}W_{dh}q_{dh}\bar{x}_{dh}^{(t)}\right)+\hat{\beta}_2\left(S_{xd}^{2(t)}-\sum_{h=1}^{L}W_{dh}q_{dh}s_{xdh}^{2(t)}\right) \qquad (7)$$

where

$$\hat{\beta}_1=\frac{\left(\sum_{h=1}^{L}W_{dh}q_{dh}\bar{x}_{dh}^{(t)}\bar{y}_{dh}^{*(t)}\right)\left(\sum_{h=1}^{L}W_{dh}q_{dh}s_{xdh}^{4(t)}\right)-\left(\sum_{h=1}^{L}W_{dh}q_{dh}\bar{x}_{dh}^{(t)}\bar{y}_{dh}^{*(t)}\right)\left(\sum_{h=1}^{L}W_{dh}q_{dh}\bar{x}_{dh}^{(t)}s_{xdh}^{2(t)}\right)}{\left(\sum_{h=1}^{L}W_{dh}q_{dh}\bar{x}_{dh}^{2(t)}\right)\left(\sum_{h=1}^{L}W_{dh}q_{dh}s_{xdh}^{4(t)}\right)-\left(\sum_{h=1}^{L}W_{dh}q_{dh}\bar{x}_{dh}^{(t)}s_{xdh}^{2(t)}\right)^2}$$

$$\boldsymbol{\hat{\beta}_2=\frac{\left(\sum_{h=1}^{L}W_{dh}q_{dh}\,\bar{x}_{dh}^{2(t)}\right)\left(\sum_{h=1}^{L}W_{dh}q_{dh}\bar{y}_{dh}^{*(t)}s_{xdh}^{2(t)}\right)-\left(\sum_{h=1}^{L}W_{dh}q_{dh}\bar{x}_{dh}^{(t)}s_{xdh}^{2(t)}\right)\left(\sum_{h=1}^{L}W_{dh}q_{dh}\bar{y}_{dh}^{*(t)}s_{xdh}^{2(t)}\right)}{\left(\sum_{h=1}^{L}W_{dh}q_{dh}\,\bar{x}_{dh}^{2}\right)\left(\sum_{h=1}^{L}W_{dh}q_{dh}\,s_{xdh}^{4}\right)-\left(\sum_{h=1}^{L}W_{dh}q_{dh}\bar{x}_{dh}s_{xdh}^{2}\right)^2}}$$

**Bias and MSE of the Proposed Estimator**

Using large sample approximations,

let $e_0^*=\frac{\bar{y}_{dh}^{*(t)}-\bar{Y}_{dh}^{(t)}}{\bar{Y}_{dh}^{(t)}}$, $e_1=\frac{\bar{x}_{dh}^{(t)}-\bar{X}_{dh}^{(t)}}{\bar{X}_{dh}^{(t)}}$, $e_2=\frac{s_{xdh}^{2(t)}-S_{dh}^{2(t)}}{S_{dh}^{2(t)}}$, $e_3=\frac{\hat{\beta}_1-\beta_1}{\beta_1}$, and $e_4=\frac{\hat{\beta}_2-\beta_2}{\beta_2}$

where $\bar{y}_{dh}^{*(t)}=\frac{n_{dh1}\bar{y}_{d1}+n_{dh2}\bar{y}_{d2}}{n_{dh}}$, $\bar{x}_{dh}^{(t)}=\frac{1}{n_{dh}}\sum_{i=1}^{n_{dh}}x_{dhi}^{(t)}$, $\bar{X}_{dh}=\frac{1}{N_{dh}}\sum_{i=1}^{N_{dh}}X_{dhi}^{(t)}$,
$s_{xdh}^2=\frac{1}{n_{dh}-1}\sum_{i=1}(x_{dhi}-\bar{x}_{dh})^2$, $\bar{X}_d=\frac{1}{N_d}\sum_{h=1}^{L}\bar{X}_{dh}$, $S_{xdh}^2=\frac{1}{N_{dh}-1}\sum_{i=1}(X_{dhi}-\bar{X}_{dh})^2$,
$S_d^2=\frac{1}{N_d-1}\sum_{h=1}(X_{dh}-\bar{X}_d)^2$

The above expressions can be summarized as follows

$$\begin{cases}\bar{y}_{dh}^{*(t)}=\bar{Y}_{dh}^{(t)}(1+e_0^*)\\ \bar{x}_{dh}^{(t)}=\bar{X}_{dh}^{(t)}(1+e_1)\\ s_{xdh}^{2(t)}=S_{dh}^{2(t)}(1+e_2)\\ \hat{\beta}_1=\beta_1(1+e_3)\\ \hat{\beta}_2=\beta_2(1+e_4)\end{cases} \qquad (8)$$

Again, let $\mathrm{E}[e_0^{*2}]=\frac{E\left[\bar{y}_{dh}^{*(t)}-\bar{Y}_{dh}^{(t)}\right]^2}{\bar{Y}_{dh}^{2(t)}}=\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)\frac{S_{ydh}^{2(t)}}{\bar{Y}_{dh}^{2(t)}}+\frac{K_{dh}-1}{\bar{Y}_{dh}^{2(t)}n_{dh2}}W_{dh2}S_{ydh2}^{2(t)}$

$\mathrm{E}[e_1^2]=\frac{E\left[\bar{x}_{dh}^{(t)}-\bar{X}_{dh}^{(t)}\right]^2}{\bar{X}_{dh}^{2(t)}}=\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)\frac{S_{x_{dh}}^{2(t)}}{\bar{X}_{dh}^{2(t)}}$, $\mathrm{E}[e_2^2]=\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)(\lambda_{04}-1)$

$E[e_0^*e_1]=E\left[\left(\frac{\bar{x}_{dh}^{(t)}-\bar{X}_{dh}^{(t)}}{\bar{X}_{dh}^{(t)}}\right)\left(\frac{\bar{y}_{dh}^{*(t)}-\bar{Y}_{dh}^{(t)}}{\bar{Y}_{dh}^{(t)}}\right)\right]=\frac{cov\left(\bar{x}_{dh}^{(t)},\bar{y}_{dh}^{*(t)}\right)}{\bar{X}_{dh}^{(t)}\bar{Y}_{dh}^{(t)}}=\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)\frac{\rho_{xy}S_{xdh}^{(t)}S_{ydh}^{(t)}}{\bar{X}_{dh}^{(t)}\bar{Y}_{dh}^{(t)}}$

where $E[e_0^*]=E\left[\frac{\bar{y}_{dh}^{*(t)}-\bar{Y}_{dh}^{(t)}}{\bar{Y}_{dh}^{(t)}}\right]=\frac{\bar{Y}_d^{(t)}-\bar{Y}_d^{(t)}}{\bar{Y}_d^{(t)}}=0$. Similarly, $E[e_1]=E[e_2]=E[e_3]=E[e_4]=0$

$E[e_1e_2] = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)\frac{S_{xdh}^{(t)}}{\bar{X}_{dh}^{(t)}}\lambda_{03}$

where

$\lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{r/2}\mu_{02}^{r/2}}$, and $\mu_{rs} = \frac{1}{N_d-1}\sum_{h=1}^{H}\left(\bar{Y}_{dh}^{(t)} - \bar{Y}_{d}^{(t)}\right)^r\left(\bar{X}_{dh}^{(t)} - \bar{X}_{d}^{(t)}\right)^s$, $\mu_{20} = S_{ydh}^{2(t)}$, $\mu_{02} = S_{xdh}^{2(t)}$

hence, $\lambda_{03} = \frac{\mu_{03}}{\mu_{20}^{0/2}\mu_{02}^{3/2}}$, and $E[e_0^*e_2] = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)\frac{S_{ydh}^{(t)}}{\bar{Y}_{dh}^{(t)}}\lambda_{12}$

To obtain bias, (8) is substituted into (7) as follows

$$\hat{\bar{y}}_{st1}^{(t)} = \sum_{h=1}^{L}W_{dh}\bar{Y}_{dh}^{(t)}(1+e_0^*) + \beta_1(1+e_3)\left(\bar{X}_d^{(t)} - \sum_{h=1}^{L}W_{dh}\bar{X}_{dh}^{(t)}(1+e_1)\right) + \beta_2(1+e_4)\left(S_d^{2(t)} - \sum_{h=1}^{L}W_{dh}S_{xdh}^{2(t)}(1+e_2)\right)$$

$$= \bar{Y}_d^{*(t)} + \sum_{h=1}^{L}W_{dh}\bar{Y}_{dh}^{(t)}e_0^* + (\beta_1 + \beta_1e_3)\left(\bar{X}_d^{(t)} - \bar{X}_d^{(t)} - \sum_{h=1}^{L}W_{dh}\bar{X}_{dh}^{(t)}e_1\right) + (\beta_2 + \beta_2e_4)\left(S_d^{2(t)} - S_d^{2(t)} - \sum_{h=1}^{L}W_{dh}S_{dh}^{2(t)}e_2\right)$$

$$= \bar{Y}_d^{(t)} + \sum_{h=1}^{L}W_{dh}\bar{Y}_{dh}^{(t)}e_0^* - \beta_1\sum_{h=1}^{L}W_{dh}\bar{X}_{dh}^{(t)}e_1 - \beta_1\sum_{h=1}^{L}W_{dh}\bar{X}_{dh}^{(t)}e_1e_3 - \beta_2\sum_{h=1}^{L}W_{dh}S_{dh}^{2(t)}e_2 - \beta_2\sum_{h=1}^{L}W_{dh}S_{dh}^{2(t)}e_2\,e_4 \quad (9)$$

Taking expectation of (9)

$$E(\hat{\bar{y}}_{st1}^{(t)}) = \bar{Y}_d^{(t)} + \sum_{h=1}^{L}W_{dh}\bar{Y}_{dh}^{(t)}E(e_0^*) - \beta_1\sum_{h=1}^{L}W_{dh}\bar{X}_{dh}^{(t)}E(e_1) - \beta_1\sum_{h=1}^{L}W_{dh}\bar{X}_{dh}^{(t)}E(e_1e_3) - \beta_2\sum_{h=1}^{L}W_{dh}S_{dh}^{2(t)}E(e_2) - \beta_2\sum_{h=1}^{L}W_{dh}S_{dh}^{2(t)}E(e_2\,e_4)$$

$$= \bar{Y}_d^{(t)} - \beta_1\sum_{h=1}^{L}W_{dh}\bar{X}_{dh}^{(t)}E(e_1e_3) - \beta_2\sum_{h=1}^{L}W_{dh}S_{dh}^{2(t)}E(e_2\,e_4)$$

$$= \bar{Y}_d^{(t)} - \beta_1\bar{X}_d^{(t)}E(e_1e_3) - \beta_2S_d^{2(t)}E(e_2e_4)$$

$$= \bar{Y}_d^{(t)} - Cov\left[\bar{x}_d^{(t)}, \hat{\beta}_1\right] - Cov\left[s_d^{2(t)}, \hat{\beta}_2\right]$$

Thus,

$$\text{Bias}\left(\hat{\bar{y}}_{st1}^{(t)}\right) = -Cov\left[\bar{x}_d^{(t)}, \hat{\beta}_1\right] - Cov\left[s_d^{2(t)}, \hat{\beta}_2\right] \quad (10)$$

To obtain MSE, we consider the following

$$\text{MSE}\left(\hat{\bar{y}}_{st1}^{(t)}\right) = \text{E}\left[\hat{\bar{y}}_{st1}^{(t)} - \bar{Y}_d^{(t)}\right]^2$$

substituting for $\hat{\bar{y}}_{st1}^{(t)}$, gives

$$= E\left[\sum_{h=1}^{L}W_{dh}\bar{Y}_{dh}^{(t)}(1+e_0^*) + \beta_1(1+e_3)\left(\bar{X}_d^{(t)} - \sum_{h=1}^{L}W_{dh}\bar{X}_{dh}^{(t)}(1+e_1)\right) + \beta_2(1+e_4)\left(S_d^{2(t)} - \sum_{h=1}^{L}W_{dh}S_{xdh}^{2(t)}(1+e_2)\right) - \bar{Y}_d^{(t)}\right]^2$$

$$= \sum_{h=1}^{L}W_{dh}^2\left[\left(\frac{1}{\mathrm{n_{dh}}} - \frac{1}{\mathrm{N_{dh}}}\right)S_{ydh}^{2(t)} + \frac{(\mathrm{k_{dh}}-1)W_{dh2}S_{ydh2}^{2(t)}}{\mathrm{n_{dh2}}}\right] - 2\beta_1\sum_{h=1}^{L}W_{dh}^2\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)\rho_{xy}S_{xdh}^{(t)}S_{ydh}^{(t)} - 2\,\beta_2\sum_{h=1}^{L}W_{dh}^2S_d^{2(t)}\left(\frac{1}{n_{dh}} - \frac{1}{N_d}\right)S_{ydh}^{(t)}\lambda_{12} + 2\beta_1\beta_2\sum_{h=1}^{L}W_{dh}^2S_{dh}^{2(t)}\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S_{xdh}^{(t)}\lambda_{03} + \beta_1^2\sum_{h=1}^{L}W_{dh}^2\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S_{xdh}^{2(t)} + \beta_2^2\sum_{h=1}^{L}W_{dh}^2\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S_{dh}^{4(t)}(\lambda_{04} - 1) \quad (11)$$

## Minimum Mean Square Error

To obtain the minimum MSE, Eq. 11 is differentiated partially with respect to $\beta_1$ and $\beta_2$

$$\beta_1 \sum_{h=1}^{L} W_{dh}^2 \left(\frac{1}{\mathrm{n_{dh}}} - \frac{1}{\mathrm{N_{dh}}}\right) S_{xdh}^{2(t)} + \beta_2 \sum_{h=1}^{L} W_{dh}^2 S_{dh}^{2(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{(t)} \lambda_{03} = \sum_{h=1}^{L} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) \rho_{xy} S_{xdh}^{(t)} S_{ydh}^{(t)}$$

Also,

$$\beta_1 \sum_{h=1}^{L} W_{dh}^2 S_{dh}^{2(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{(t)} \lambda_{03} + \beta_2 \sum_{h=1}^{L} W_{dh}^2 S_{dh}^{4(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) (\lambda_{04} - 1) = \sum_{h=1}^{L} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{ydh}^{3(t)} \lambda_{12}$$

solving for $\beta_1$ and $\beta_2$ simultaneously give

$$\beta_{1opt} = \frac{\left(\sum_{h=1}^{L} W_{dh}^2\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)\rho_{xy} S_{xdh}^{(t)} S_{ydh}^{(t)}\right)\left(\sum_{h=1}^{L} W_{dh}^2 S_{dh}^{4(t)}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)(\lambda_{04}-1)\right) - \left(\sum_{h=1}^{L} W_{dh}^2\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right) S_{ydh}^{3(t)}\lambda_{12}\right)\left(\sum_{h=1}^{L} W_{dh}^2 S_{dh}^{2(t)}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right) S_{xdh}^{(t)}\lambda_{03}\right)}{\left(\sum_{h=1}^{L} W_{dh}^2\left(\frac{1}{\mathrm{n_{dh}}}-\frac{1}{\mathrm{N_{dh}}}\right) S_{xdh}^{2(t)}\right)\left(\sum_{h=1}^{L} W_{dh}^2 S_{dh}^{4(t)}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)(\lambda_{04}-1)\right) - \left(\sum_{h=1}^{L} W_{dh}^2 S_{dh}^{2(t)}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right) S_{xdh}^{(t)}\lambda_{03}\right)^2} \quad (12)$$

$$\beta_{2opt} = \frac{\left(\sum_{h=1}^{L} W_{dh}^2\left(\frac{1}{\mathrm{n_{dh}}}-\frac{1}{\mathrm{N_{dh}}}\right) S_{xdh}^{2(t)}\right)\left(\sum_{h=1}^{L} W_{dh}^2\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right) S_{ydh}^{3(t)}\lambda_{12}\right) - \left(\sum_{h=1}^{L} W_{dh}^2 S_{dh}^{2(t)}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right) S_{xdh}^{(t)}\lambda_{03}\right)\left(\sum_{h=1}^{L} W_{dh}^2\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)\rho_{xy} S_{xdh}^{(t)} S_{ydh}^{(t)}\right)}{\left(\sum_{h=1}^{L} W_{dh}^2\left(\frac{1}{\mathrm{n_{dh}}}-\frac{1}{\mathrm{N_{dh}}}\right) S_{xdh}^{2(t)}\right)\left(\sum_{h=1}^{L} W_{dh}^2 S_{dh}^{4(t)}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right)(\lambda_{04}-1)\right) - \left(\sum_{h=1}^{L} W_{dh}^2 S_{dh}^{2(t)}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right) S_{xdh}^{(t)}\lambda_{03}\right)^2} \quad (13)$$

Substituting for Eqs. 12 and 13 in Eq.11 gives

$$\mathrm{MSE}_{\min}\left(\hat{\bar{y}}_{st1}^{(t)}\right) = \sum_{h=1}^{L} W_{dh}^2 \left[\left(\frac{1}{\mathrm{n_{dh}}} - \frac{1}{\mathrm{N_{dh}}}\right) S_{ydh}^{2(t)} + \frac{(\mathrm{k_{dh}}-1) W_{dh2} S_{ydh2}^{2(t)}}{\mathrm{n_{dh2}}}\right] - 2\beta_{1opt} \sum_{h=1}^{L} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) \rho_{xy} S_{xdh}^{(t)} S_{ydh}^{(t)} - 2\,\beta_{2opt} \sum_{h=1}^{L} W_{dh}^2 S_{d}^{2(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{d}}\right) S_{ydh}^{(t)} \lambda_{12} + 2\beta_{1opi}\beta_{2opi} \sum_{h=1}^{L} W_{dh}^2 S_{dh}^{2(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{(t)} \lambda_{03} + \beta_{1opt}^2 \sum_{h=1}^{L} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{2(t)} + \beta_{2opt}^2 \sum_{h=1}^{L} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{dh}^{4(t)} (\lambda_{04} - 1) \quad (14)$$

**Case II:** By adopting condition 2, the properties of the proposed estimator when nonresponse is observed in both the study and the auxiliary variables are as follows

Let $\bar{x}_{dh}^{*(t)}$ be analogously defined as $\bar{y}_{dh}^{*(t)}$ with sub-sampling the nonresponse strategy, thus

$$\bar{x}_{dh}^{*(t)} = \frac{n_{dh1}\bar{x}_{d1} + n_{dh2}\bar{x}_{d2}}{n_{dh}}$$

Then Eq. 6 can be written as

$$\hat{\bar{y}}_{st2}^{(t)} = \sum_{h=1}^{L} W_{dh} q_{dh}\, \bar{y}_{dh}^{*(t)} + \hat{\beta}_1 \left(\bar{X}_{d}^{(t)} - \sum_{h=1}^{L} W_{dh} q_{dh} \bar{x}_{dh}^{*(t)}\right) + \hat{\beta}_2 (S_{xdh}^{2(t)} - \sum_{h=1}^{L} W_{dh} q_{dh} s_{xdh}^{2(t)}) \quad (15)$$

Such that $\bar{x}_{dh}^{*(t)} = \bar{X}_{dh}^{(t)} (1 + e_1^*)$ and

$$\begin{cases} \mathrm{E}[e_1^{2*}] = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)\frac{S_{xdh}^{2(t)}}{\bar{X}_{dh}^{2(t)}} + \frac{K_{dh}-1}{\bar{X}_{dh}^{2(t)}\mathrm{n}_{\mathrm{dh2}}} W_{dh2}S_{xdh2}^{2(t)} \\ E(e_0^* e_1^*) = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)\rho_{xy_{dh}}\frac{S_{xdh}^{(t)}}{\bar{X}_{dh}^{(t)}}\frac{S_{ydh}^{(t)}}{\bar{Y}_{dh}^{(t)}} + \frac{K_{dh}-1}{\bar{X}_{dh}^{(t)}\bar{Y}_{dh}^{(t)}\mathrm{n}_{\mathrm{dh2}}} W_{dh2}\rho_{xy_{dh2}}S_{xdh2}^{(t)}S_{ydh2}^{(t)} \\ E[e_1^* e_2] = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)\frac{S_{xdh}^{(t)}}{\bar{X}_{dh}^{(t)}}\lambda_{03} \end{cases}$$

Then Eq. 15 can be written in terms of the large sample approximations as

$$\hat{\bar{y}}_{st2}^{(t)} = \sum_{h=1}^{L} W_{dh}\bar{Y}_{dh}^{*(t)}(1+e_0^*) + \beta_1(1+e_3)\left(\bar{X}_d^{(t)} - \sum_{h=1}^{L} W_{dh}\bar{X}_{dh}^{(t)}(1+e_1^*)\right) + \beta_2(1+e_4)\left(S_d^{2(t)} - \sum_{h=1}^{L} W_{dh}S_{xdh}^{2(t)}(1+e_2)\right)$$

$$= \bar{Y}_d^{*(t)} + \sum_{h=1}^{L} W_{dh}\bar{Y}_{dh}^{*(t)}e_0^* - \beta_1\sum_{h=1}^{L} W_{dh}\bar{X}_{dh}^{(t)}e_1^* - \beta_1\sum_{h=1}^{L} W_{dh}\bar{X}_{dh}^{(t)}e_1^* e_3 - \beta_2\sum_{h=1}^{L} W_{dh}S_{dh}^{2(t)}e_2 - \beta_2\sum_{h=1}^{L} W_{dh}S_{dh}^{2(t)}e_2\, e_4 \qquad (16)$$

Taking expectation of Eq. 16 and substituting for the large sample approximations

$$E(\hat{\bar{y}}_{st2}^{(t)}) = \bar{Y}_d^{*(t)} - \sum_{h=1}^{L} W_{dh}\bar{X}_{dh}^{(t)}E(e_1^* e_3) - \beta_2\sum_{h=1}^{L} W_{dh}S_{dh}^{2(t)}E(e_2\, e_4)$$

$$= \bar{Y}_d^{*(t)} - \beta_1\bar{X}_d^{(t)}E(e_1^* e_3) - \beta_2 S_d^{2(t)}E(e_2 e_4)$$

$$= \bar{Y}_d^{*(t)} = -Cov\left[\bar{x}_d^{(t)}, \hat{\beta}_1\right] - Cov\left[s_d^{2(t)}, \hat{\beta}_2\right]$$

$$\mathrm{Bias}\left(\hat{\bar{y}}_{st2}^{(t)}\right) = -Cov\left[\bar{x}_d^{(t)}, \hat{\beta}_1\right] - Cov\left[s_d^{2(t)}, \hat{\beta}_2\right] \qquad (17)$$

***MSE of the proposed estimator***

**Proof:**

Let the MSE of $\hat{\bar{y}}_{st2}^{(t)}$ under the condition 2 be denoted as MSE $\left(\hat{\bar{y}}_{st2}^{(t)}\right)$

$$\mathrm{MSE}\left(\hat{\bar{y}}_{st2}^{(t)}\right) = \mathrm{E}\left[\hat{\bar{y}}_{st2}^{(t)} - \bar{Y}_d^{(t)}\right]^2$$

substituting for $\hat{\bar{y}}_{st2}^{(t)}$ gives

$$= E\left[\sum_{h=1}^{L} W_{dh}\bar{Y}_{dh}^{(t)}(1+e_0^*) + \beta_1(1+e_3)\left(\bar{X}_d^{(t)} - \sum_{h=1}^{L} W_{dh}\bar{X}_{dh}^{(t)}(1+e_1^*)\right) + \beta_2(1+e_4)\left(S_d^{2(t)} - \sum_{h=1}^{L} W_{dh}S_{xdh}^{2(t)}(1+e_2)\right) - \bar{Y}_d^{(t)}\right]^2$$

$$MSE(\hat{\bar{y}}_{st2}^{(t)}) = \sum_{h=1}^{L} W_{dh}^2\,\bar{Y}_{dh}^{2(t)}E(e_0^{2*}) - 2\beta_1\sum_{h=1}^{L} W_{dh}^2\bar{X}_{dh}^{(t)}\bar{Y}_{dh}^{(t)}\,E(e_0^* e_1^*) - 2\,\beta_2\sum_{h=1}^{L} W_{dh}^2\bar{Y}_{dh}^{(t)}S_{dh}^{2(t)}\,E(e_0^* e_2) + 2\beta_1\beta_2\sum_{h=1}^{L} W_{dh}^2\bar{X}_{dh}^{(t)}S_{dh}^{2(t)}\,E(e_1^* e_2) + \beta_1^2\sum_{h=1}^{L} W_{dh}^2\,\bar{X}_{dh}^{2(t)}E(e_1^{2*}) + \beta_2^2\sum_{h=1}^{L} W_{dh}^2\,S_{dh}^{4(t)}E(e_2^2)$$

$$= \sum_{h=1}^{L} W_{dh}^2\left[\left(\frac{1}{\mathrm{n_{dh}}} - \frac{1}{\mathrm{N_{dh}}}\right)S_{ydh}^{2(t)} + \frac{(\mathrm{k_{dh}}-1)W_{dh2}S_{ydh2}^{2(t)}}{\mathrm{n_{dh2}}}\right] - 2\beta_1\sum_{h=1}^{L} W_{dh}^2\left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)\rho_{xy_{dh}}S_{xdh2}^{(t)}S_{ydh2}^{(t)} + \frac{K_{dh}-1}{n_{dh2}}W_{dh2}\rho_{xy_{dh2}}S_{xdh2}^{(t)}S_{ydh2}^{(t)}\right] - 2\,\beta_2\sum_{h=1}^{L} W_{dh}^2 S_d^{2(t)}\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S_{ydh}^{2(t)}\lambda_{12} + 2\beta_1\beta_2\sum_{h=1}^{L} W_{dh}^2 S_d^{2(t)}\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S_{xdh}^{(t)}\lambda_{03} + \beta_1^2\sum_{h=1}^{L} W_{dh}^2\left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S_{xdh}^{2(t)} + \frac{K_{dh}-1}{n_{dh2}}W_{dh2}S_{xdh2}^{2(t)}\right] + \beta_2^2\sum_{h=1}^{L} W_{dh}^2\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S_{dh}^{4(t)}(\lambda_{04} - 1) \qquad (18)$$

To obtain minimum MSE, we differentiate (18) partially with respect to $\beta_1 and\ \beta_2$

$$\beta_1 \sum_{h=1}^{L} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{2(t)} + \frac{K_{dh}-1}{n_{dh2}} W_{dh2} S_{xdh2}^{2(t)}\right] + \beta_2 \sum_{h=1}^{L} W_{dh}^2 S_{dh}^{2(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{(t)} \lambda_{03} = \sum_{h=1}^{L} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) \rho_{xy_{dh}} S_{xdh2}^{(t)} S_{ydh2}^{(t)} + \frac{K_{dh}-1}{n_{dh2}} W_{dh2} \rho_{xy_{dh2}} S_{xdh2}^{(t)} S_{ydh2}^{(t)}\right]$$

$$\beta_1 \sum_{h=1}^{L} W_{dh}^2 S_{dh}^{2(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{(t)} \lambda_{03} + \beta_2 \sum_{h=1}^{L} W_{dh}^2\, S_{dh}^{4(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) (\lambda_{04} - 1) \ = \sum_{h=1}^{L} W_{dh}^2 S_{d}^{2(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{ydh}^{2(t)} \lambda_{12}$$

solving for $\beta_1$ and $\beta_2$ simultaneously give

$$\beta_{1OPT} = \frac{\left(\sum_{h=1}^{L} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) \rho_{xy_{dh}} S_{xdh2}^{(t)} S_{ydh2}^{(t)} + \frac{K_{dh}-1}{n_{dh2}} W_{dh2} \rho_{xy_{dh2}} S_{xdh2}^{(t)} S_{ydh2}^{(t)}\right]\right)\left(\sum_{h=1}^{L} W_{dh}^2 S_{dh}^{4(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)(\lambda_{04}-1)\right) - \left(\sum_{h=1}^{L} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{ydh}^{3(t)} \lambda_{12}\right)\left(\sum_{h=1}^{L} W_{dh}^2 S_{dh}^{2(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{(t)} \lambda_{03}\right)}{\left(\sum_{h=1}^{L} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{2(t)} + \frac{K_{dh}-1}{n_{dh2}} W_{dh2} S_{xdh2}^{2(t)}\right]\right)\left(\sum_{h=1}^{L} W_{dh}^2 S_{dh}^{4(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)(\lambda_{04}-1)\right) - \left(\sum_{h=1}^{L} W_{dh}^2 S_{dh}^{2(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{(t)} \lambda_{03}\right)^2}$$

$$\beta_{2OPT} = \frac{\left(\sum_{h=1}^{L} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{2(t)} + \frac{K_{dh}-1}{n_{dh2}} W_{dh2} S_{xdh2}^{2(t)}\right]\right)\left(\sum_{h=1}^{L} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{ydh}^{3(t)} \lambda_{12}\right) - \left(\sum_{h=1}^{L} W_{dh}^2 S_{dh}^{2(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{(t)} \lambda_{03}\right)\left(\sum_{h=1}^{L} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) \rho_{xy_{dh}} S_{xdh2}^{(t)} S_{ydh2}^{(t)} + \frac{K_{dh}-1}{n_{dh2}} W_{dh2} \rho_{xy_{dh2}} S_{xdh2}^{(t)} S_{ydh2}^{(t)}\right]\right)}{\left(\sum_{h=1}^{L} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{2(t)} + \frac{K_{dh}-1}{n_{dh2}} W_{dh2} S_{xdh2}^{2(t)}\right]\right)\left(\sum_{h=1}^{L} W_{dh}^2 S_{dh}^{4(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)(\lambda_{04}-1)\right) - \left(\sum_{h=1}^{L} W_{dh}^2 S_{dh}^{2(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{(t)} \lambda_{03}\right)^2}$$

Substituting for $\beta_{1OPT}$ and $\beta_{2OPT}$ in (18), we have

$$\mathrm{MSE}_{\min}\left(\hat{\hat{\bar{y}}}_{st2}^{(t)}\right) = \sum_{h=1}^{L} W_{dh}^2 \left[\left(\frac{1}{\mathrm{n_{dh}}} - \frac{1}{\mathrm{N_{dh}}}\right) S_{ydh}^{2(t)} + \frac{(\mathrm{k_{dh}}-1) W_{dh2} S_{ydh2}^{2(t)}}{\mathrm{n_{dh2}}}\right] - 2\beta_{1OPT} \sum_{h=1}^{L} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) \rho_{xy_{dh}} S_{xdh2}^{(t)} S_{ydh2}^{(t)} + \frac{K_{dh}-1}{n_{dh2}} W_{dh2} \rho_{xy_{dh2}} S_{xdh2}^{(t)} S_{ydh2}^{(t)}\right] - 2\,\beta_{2OPT} \sum_{h=1}^{L} W_{dh}^2 S_{d}^{2(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{ydh}^{2(t)} \lambda_{12} + 2\beta_{1OPT}\beta_{2OPT} \sum_{h=1}^{L} W_{dh}^2 S_{d}^{2(t)} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{(t)} \lambda_{03} + \beta_{1OPT}^2 \sum_{h=1}^{L} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{2(t)} + \frac{K_{dh}-1}{n_{dh2}} W_{dh2} S_{xdh2}^{2(t)}\right] + \beta_{2OPT}^2 \sum_{h=1}^{L} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{dh}^{4(t)} (\lambda_{04} - 1) \qquad (19)$$

## 3. Empirical Study

A real-life data obtained from the Household Finances and Consumption Survey (HFCS) and the Integrated Household Survey (IHS) from the Statistics Department of Central Bank of Nigeria which is a quarterly survey is partitioned into WAVE 1,2, 3 & 4 for 2019 (before COVID-19), WAVE 1, 2, 3 & 4 for 2020 (during COVID-19) and WAVE 2, 3 & 4 for 2021 (After COVID-19). The population is comprised of selected households spread across the 37 states (domains) of Nigeria. Again, the population is sub-divided into five groups (strata) as follows: those working in government organizations, private organizations, self-employed with own employees, self-employed without own employees, and unpaid family workers.

### 3.1 Performance of Estimators

The performance of the existing and proposed estimators across the domains are presented in Tables 1, 2 and 3 using average variance and mean square errors of the respective estimators.

**Table 1: Performance of estimators across the domains using average variance and MSE in 2019**

| ESTIMATORS | WAVE 1 | WAVE 2 | WAVE 3 | WAVE 4 |
|---|---|---|---|---|
| $V(\hat{\bar{y}}_d)$ | 22352872 | 11278590 | 2127110381 | 2128536792 |
| $MSE(T_{DR.st.d})$ | 198548137 | 437916767 | 300383822 | 334666336 |
| $MSE(T_{DG.st.D})$ | 55781052 | 435788250 | 433161749 | 4277954577 |
| $MSE\left(T_{DG.st.\beta.d}\right)$ | 45537277 | 41154765 | 31484947 | 34695256 |
| $MSE(t_{cal}^{*})_{min}$ | 33893634 | 416275960 | 48609844 | 3208332609 |
| $\mathrm{MSE}_{\min}(\hat{\bar{y}}_{st1}^{(t)})$ | 19334 | 61544 | 51873 | 411090 |
| $\mathrm{MSE}_{\min}(\hat{\bar{y}}_{st2}^{(t)})$ | 8321294 | 9801300 | 822347 | 875463 |

Source: Authors' computation 2026

**Table 2: Performance of estimators across the domains using average variance and MSE in 2020**

| ESTIMATORS | WAVE 1 | WAVE 2 | WAVE 3 | WAVE 4 |
|---|---|---|---|---|
| $V(\hat{\bar{y}}_d)$ | 12070397032 | 1959390531 | 2696529563 | 2877211507 |
| $MSE(T_{DR.st.d})$ | 3174210324 | 6566261196 | 1891139700 | 2550790226 |
| $MSE(T_{DG.st.D})$ | 874122793 | 945289242 | 42607369 | 32456723 |
| $MSE\left(T_{DG.st.\beta.d}\right)$ | 764959433 | 45621109 | 411234543 | 354960584 |
| $MSE(t_{cal}^{*})_{min}$ | 40122897 | 576008934 | 11803955 | 10912063 |
| $\mathrm{MSE}_{\min}(\hat{\bar{y}}_{st1}^{(t)})$ | 66651 | 456395 | 47228 | 55217 |
| $\mathrm{MSE}_{\min}(\hat{\bar{y}}_{st2}^{(t)})$ | 343673 | 2296842 | 436733 | 212321 |

Source: Authors' computation 2026

**Table 3: Performance of estimators across the domains using average variance and MSE in 2021**

| ESTIMATORS | WAVE 2 | WAVE 3 | WAVE 4 |
|---|---|---|---|
| $V(\hat{\bar{y}}_d)$ | 13079119911 | 8091356206 | 8261505450 |
| $MSE(T_{DR.st.d})$ | 3529448053 | 6219274382 | 9999628351 |
| $MSE(T_{DG.st.D})$ | 5083972010 | 5963101027 | 454438216 |
| $MSE\left(T_{DG.st.\beta.d}\right)$ | 634849473 | 534859495 | 412455644 |
| $MSE(t_{cal}^{*})_{min}$ | 9062483839 | 4782374466 | 55175506557 |
| $\mathrm{MSE}_{\min}(\hat{\bar{y}}_{st1}^{(t)})$ | 23453 | 21145 | 34647 |
| $\mathrm{MSE}_{\min}(\hat{\bar{y}}_{st2}^{(t)})$ | 223465 | 43561 | 24683431 |

Source: Authors' computation 2026

## 4. Discussion

This discussion is based on the empirical analysis carried out and presented in Tables 1, 2 and 3. From the result of analysis in Tables 1, 2 and 3, it is observed that the average minimum mean square error of the proposed estimators $\hat{\bar{y}}_{st1}^{(t)}$ and $\hat{\bar{y}}_{st2}^{(t)}$ are far less than the average Variance and MSEs of the existing estimators $\hat{\bar{y}}_d, T_{DR.st.d}, T_{DG.st.D}, T_{DG.st,\beta.d}$ and $t_{cal}^{*}$ across all the domains in all the waves of 2019, 2020 and 2021. This agrees with the literature in Bassey and Iseh (2025) on the use of the Kullback-Leibler distance measure for weight adjustment in small area estimation. In addition, it is observed that the estimator $\hat{\bar{y}}_{st1}^{(t)}$ has performed more efficiently than $\hat{\bar{y}}_{st2}^{(t)}$, probably because nonresponse affected both the study and auxiliary variables of the estimator $\hat{\bar{y}}_{st2}^{(t)}$.

## 5. Conclusion

This study addresses the challenges facing small area estimation (SAE), particularly in the presence of nonresponse. By extending the Hansen & Hurwitz (1946) estimator, two new calibration-based estimators are proposed using the Kullback-Leibler distance function-based weight adjustment to improve efficiency. The proposed estimators are formulated under two conditions: when nonresponse affects only the study variable and when nonresponse affects both the study and auxiliary variables.

The empirical results show that the suggested estimator improves efficiency and minimizes estimation errors when compared to current estimators. Empirical investigations using real-life data from the HFCS and IHS from the Central Bank of Nigeria's Statistics Department (2019-2021) to validate the theoretical derivations have revealed that the proposed estimators outperform existing estimators in terms of efficiency gains. This illustrates that the suggested estimators have asymptotic features that beat existing SAE estimators during various survey periods (before, during, and after COVID-19).

The empirical investigation found that subsampling the nonresponse using calibration approaches considerably improved domain mean estimate in stratified sampling. The suggested estimators give a robust and computationally feasible strategy for dealing with small sample sizes and missing data, which can result in over/under estimation or sometimes yield negative weights in small area estimation..

## References

Anekeya, A. D., Onyango, O. C., & Nyongesa, K. (2018). Domain mean estimation using double sampling with non-linear cost function in the non-response presence. Science Journal of Applied Mathematics and Statistics, 6(1), 42. https://doi.org/10.11648/j.sjams.20180601.14

Ashutosh. Estimator of Domain Mean Using Stratified Sampling in the Presence of Non-Response, *Sri Lankan Journal of Applied Statistics*, 2021; DOI:10.4038/sljastats.v22i1.8049

Bassey, M. O. and Iseh, M. J. Two-phase sampling with calibration weight Adjustment in small-area area estimation for Longitudinal surveys. *International Journal of nterdisciplinary Research in Statistics, Mathematics and Engineering*, 2025; 12(3):1-14. DOI: https://doi.org/10.5281/zenodo.16366887

Datta, G. S. Model-based approach to small area estimation In Sample Surveys: Inference andAnalysis, (D. Pfeffermann and C. R. Rao, eds.). Handbook of Statistics 29B, 2009; 251– 288. North-Holland, Amsterdam.

Gonzalez, M.E. Use and evaluation of synthetic estimators. Proceedings of the Social Sciences Section, Amer. Stat. Assoc.,1973;33-36.

Hansen, M. H., and Hurwitz, W. N. The problem of non-response in sample surveys. *Journal of the American Statistical Association*,1946; **41**, 517-529.

Iseh, M. J, and Bassey, K. J. A New Calibration Estimator of Population Mean for Small Area with Nonresponse. *Asian Journal of Probability and Statistics,*2021;12(2):41-51. DOI: 10.9734/AJPAS/2021/v12i230286.

Iseh, M. J and Bassey M. O. Calibration Estimators for Population Mean with Subsampling the Nonrespondents

Under Stratified Sampling. *Science Journal of Applied Mathematics and Statistics*, 2022;10(4): 45-56. DOI: *10.11648/j.sjams.20221004.11*

Iseh, M. J and Enang, E. I. A calibrated synthetic estimator for small area estimation *Statistics in Transition new series,* 2021:22(3); 15–30, DOI 10.21307/stattrans-2021-025

Iseh, M. J., Inyang, E. E. and Enang, E. I. Estimation of Population Mean in Domains of Study Under Double Sampling: A Calibration Technique for Subsampling the Nonrespondents. *Sri Lankan Journal of Applied Statistics*, 2025; 26(2):40-68. DOI: https://doi.org/10.4038/sljas.v26i2.8168.

Iseh, M. J and Bassey M. O. Smoothing of Estimators of Population mean using Calibration Technique with Sample Errors. *Journal of Modern Applied Statistical Methods*, (2024), 23(1), 1-20. https://doi.org/10.56801/Jmasm.V23.i1.2

Ikot, E. E, and Iseh, M. J. Calibration for efficiency of ratio estimator in domains of study with sub-sampling the nonrespondents. *Biom Biostat Int J.* 2024;13(2):42–50. DOI:10.15406/bbij.2024.13.00413.

Jiang, J. and Lahiri, P. Estimation of finite population domain means: A model- assisted empirical best prediction approach. J. Amer. Statist. Assoc., 2006; 101 301–311. MR2268

*Khare, B. B. ,*Ashutosh , and Rai, P. K. A Comparative Study of a Class of Direct Estimators for Domain Mean with a Direct Ratio Estimator for Domain Mean using Auxiliary Character *Statistics In Transition new series,* 2021; *22:189–200, DOI:10.21307/stattrans.*

Rao, J. N. K. *Small Area Estimation*, Wiley Inter-Science, John Wiley and Sons, New Jersey, 2003.

Särndal, C. E. "Design-consistent versus Model-dependent Estimation for Small Domains." Journal of the American Statistical Association 1984; 79 (387): 624-631. https://doi.org/10.2307/2288409.

Singh, B. and Sisodia, B.V.S. Small area estimation in longitudinal surveys. J. of Rel.and Stat. Stu., 2011; 4(2): 83-91.

Tikkiwal, G. C. and Ghiya A. A generalized class of synthetic estimators with application to crop acreage estimation for small domains. Biomet J. 2000; 42(7):865–876