EXTENSION OF INFORMATION CASES ON MIXTURE REGRESSION ESTIMATORS USING MULTI-AUXILIARY VARIABLES AND ATTRIBUTES IN TWO-PHASE SAMPLING

PETER I. OGUNYINKA and A. A. SODIPO

Department of Mathematical Sciences
Olabisi Onabanjo University
Ago-Iwoye
Nigeria
e-mail: ogunyinka.peter@gmail.com

Department of Statistics
University of Ibadan
Ibadan
Nigeria

Abstract

In survey sampling, auxiliary variables and attributes have been employed to reduce the mean square error (MSE) of estimators, hence, improving such estimators in two-phase sampling. Similarly, the three ways of utilizing auxiliary variables and attributes (full, partial, and no information cases) have provided flexibility in its usage. In this article, we have proposed two additional cases to the existing partial information case to make up five information cases of utilizing auxiliary variables and attributes. Similarly, schema of each of the proposed estimators were introduced for estimator formation. It was ascertained

2010 Mathematics Subject Classification: 62D05.
Keywords and phrases: partial information case, mixture regression estimator, two-phase sampling, auxiliary variable, auxiliary attribute.
Received December 26, 2016

© 2017 Scientific Advances Publishers
that partial information case II (PIC-II) and partial information case III (PIC-III) are efficient over partial information case I (PIC-I). Hence, the proposed two partial information cases (PIC-II and PIC-III) gained improved efficiency over the existing partial information case I (PIC-I).

1. Introduction

Amongst the survey statisticians, the use of auxiliary information has been established and been in use towards improving the estimation on the study variable. The use of auxiliary information is highly recommended when there is high correlation between the study and the auxiliary variables. Two-phase sampling, among other sampling techniques, maximizes the advantages of auxiliary variable (utilization of auxiliary information). Neyman [9, 10] initiated the use of auxiliary variable at the pre-selection stage while Watson [15] first used auxiliary variable at the post-selection stage using regression estimation method. The use of regression estimator in two-phase sampling towards estimating the study variable uses the auxiliary information at the post-selection or estimation stage. Olkin [11] pioneered the application of highly correlated multi-auxiliary variables (more than one auxiliary variable) in ratio estimation method with improved result over no auxiliary or one auxiliary variable.

Ahmad et al. [1] summarized the works of Tripathi [14] and Das [4] into four ways which auxiliary information may be available in two-phase sampling (availability of auxiliary information). Samiuddin and Hanif [13] categorize the estimators into three cases based on the availability of two or more auxiliary variables in such estimators. The three cases are full information case (FIC), no information case (NIC) and partial information case (PIC). These three cases provide the flexibility in the usage of auxiliary information depending on the various forms of availability of such auxiliary variables.

A new auxiliary feature about the population could be dichotomous property (present or absent) which is as well highly correlated with the study variable. The use of such dichotomous property, called auxiliary
attribute, in survey statistics has revealed improvement on the estimation of the study variable. Bahl and Tuteja [3], Jhajj et al. [6], Rajesh et al. [12], Hanif et al. [5], and Moeen et al. [8] are among the literatures that have utilized auxiliary attributes to obtain improve estimators over the singular use of auxiliary variable and over non-usage of auxiliary variable. Kung’u et al. [7] proposed generalized mixture regression estimator in two-phase sampling with the combination of multiple auxiliary variables and attributes following the prior three ways of utilizing the auxiliary variables, full, no and partial information cases as established by Samiuddin and Hanif [13]. The established estimators have smaller mean square errors over the prior reviewed estimators of Moeen et al. [8], hence, making them superior.

The research focuses on the partial information case established by Kung’u et al. [7], this article considers two-additional cases of such partial information case to make up three partial information cases (PIC-I, PIC-II, and PIC-III). These sum up to five information cases considering the full and no information cases. The estimator schema of the five information cases were introduced and the mean square error were established following Arora and Bansi [2] approach of presenting mean square error.

2. Preliminaries

2.1. Notations and assumptions

Considering \( N \) as the population size and \( n_1 \) and \( n_2 \) as the first and second phase sample sizes (simple random without replacement) respectively, where \( n_1 > n_2 \). Hence, presenting

\[
\theta_1 = \left( \frac{1}{n_1} - \frac{1}{N} \right) \quad \text{and} \quad \theta_2 = \left( \frac{1}{n_2} - \frac{1}{N} \right) ; \text{ for } \theta_1 < \theta_2.
\] (1)
Let $x_{(1)i}$ and $x_{(2)i}$ be the $i$-th auxiliary variable at the first and second phase sample, respectively; $y_2$ be the study variable at the second phase sampling. Then

$$
\bar{y}_2 = (\bar{Y}_y + \bar{\varepsilon}_y); \ \bar{x}_{(1)i} = (\bar{X}_i + \bar{\varepsilon}_{x(1)i}); \ \bar{x}_{(2)i} = (\bar{X}_i + \bar{\varepsilon}_{x(2)i}); \ \text{for } i = 1, 2, \ldots, p,
$$

where $\bar{\varepsilon}_y$, $\bar{\varepsilon}_{x(1)i}$, $\bar{\varepsilon}_{x(2)i}$ are the mean sampling errors and are very small, such that

$$E(\bar{\varepsilon}_{x(1)i}) = E(\bar{\varepsilon}_{x(2)i}) = 0. \tag{3}$$

Similarly, considering $\tau_{ij}$ as a complete dichotomous property about the population which is presented as

$$f(x) = \begin{cases} 1, & \text{is the } j\text{-th unit of population possessing } i\text{-th auxiliary attributes,} \\ 0, & \text{otherwise,} \end{cases} \quad j = 1, 2, \ldots, n, \tag{4}$$

$\tau_j = \text{value of } j\text{-th auxiliary attribute with the assumption that the complete dichotomy is recorded for each attribute.}$

Let $A_j = \sum_{j=1}^{N} \tau_{ij}$ and $a_j = \sum_{j=1}^{n} \tau_{ij}$ be the total number of units in the population and sample, respectively, possessing attribute $\tau_j$. Let

$$P_j = \frac{A_j}{N} \quad \text{and} \quad p_j = \frac{a_j}{n} \quad \text{be the corresponding population and sample proportion possessing attribute } \tau_j. \tag{5}$$

Similarly,

$$P_{(1)i} = (P_i + \bar{\varepsilon}_{\tau(1)i}); \ P_{(2)i} = (P_i + \bar{\varepsilon}_{\tau(2)i}),$$

for $E(\bar{\varepsilon}_{\tau(1)i}) = E(\bar{\varepsilon}_{\tau(2)i}) = 0, \tag{6}$

and

$$C_{\tau i}^2 = \frac{S_{\tau i}^2}{p^2} ; \ C_{\tau 1}^2 = \frac{S_{\tau 1}^2}{p^2} ; \ C_{\tau y}^2 = \frac{S_{\tau y}^2}{S_{\tau}^{SS}},$$

where $S_{\tau i}^2$ and $S_{\tau 1}^2$ are the variances of the $i$-th and $1$-st auxiliary variables at the second phase sampling respectively, and $S_{\tau y}^2$ is the variance of the study variable at the second phase sampling.
2.2. Mixture regression estimator in two-phase sampling

2.2.1. Full information case (FIC)

Kung’u et al. [7] established the estimated population mean of a mixture regression estimator in two-phase sampling using multi-auxiliary variables and attributes when information on all the auxiliary variables and attributes are available from the population. This is called full information case (FIC).

\[ \hat{t}_{MR(3.0)} = \bar{y}_2 + \sum_{i=1}^{k} \alpha_i (\bar{X}_i - \bar{x}_{i(2)}) + \sum_{j=k+1}^{q} \beta_j (P_j - p_{(2)}). \]  

(7)

The corresponding mean square error is given as

\[ MSE \left( \hat{t}_{MR(3.0)} \right) = 0.2 \bar{y}^2 C^2 y \left( 1 - \rho^2_{y,(\bar{x},\bar{z})_q} \right). \]

(8)

2.2.2. No information case (NIC)

The estimated population mean is established when population information on all auxiliary variables and attributes are not available. The mixture regression estimator using multi-auxiliary variables and attributes in two-phase sampling is presented by Kung’u et al. [7] as

\[ \hat{t}_{MR(3.2)} = \bar{y}_2 + \sum_{i=1}^{k} \alpha_i (\bar{X}_i - \bar{x}_{i(2)}) + \sum_{j=k+1}^{q} \beta_j (P_j - p_{(2)}). \]

(9)

The corresponding mean square error is given as

\[ MSE \left( \hat{t}_{MR(3.2)} \right) = \bar{y}^2 C^2 y \left[ \theta_2 \left( 1 - \rho^2_{y,(\bar{x},\bar{z})_q} \right) + \theta_1 \left( 1 - \rho^2_{y,(\bar{x},\bar{z})_q} \right) \right]. \]

(10)

2.2.3. Partial information case I (PIC-I)

Kung’u et al. [7] presented the estimated population mean of a mixture estimator in two-phase sampling when there are multi-auxiliary variables and attributes such that we do not have information on \( k \) auxiliary variables and \( q \) auxiliary attributes from the population.
Kung’u et al. [7] uses the second method of configuring partial information case (PIC) out of the two ways expressed by Ahmad et al. [1] of presenting partial information case. The estimator is presented as thus:

$$t^{\text{MR}(3.1)} = \left\{ \bar{y}_2 + \sum_{i=1}^{t} \alpha_i (\bar{x}_{(1)i} - \bar{x}_{(2)i}) + \sum_{i=1}^{t} \sigma_i (\bar{x}_i - \bar{x}_{(1)i}) \right\}$$

$$+ \sum_{i=t+1}^{k} \alpha_j (\bar{x}_{(1)j} - \bar{x}_{(2)j}) + \sum_{j=k+1}^{g} \beta_f (p_{(1)j} - p_{(2)f})$$

$$+ \sum_{f=k+1}^{g} \gamma_f (p_f - p_{(1)f}) + \sum_{m=g+1}^{q} \beta_m (p_{(1)m} - p_{(2)m}) \right\}. \quad (11)$$

The corresponding mean square error is given as

$$MSE (t^{\text{MR}(3.1)}) = \bar{y}^2 \chi^2 \left[ \theta_1 \left( \rho^2_{\gamma(\bar{X},\bar{Z})_q} - \rho^2_{\gamma(\bar{X},\bar{Z})_q} \right) + \theta_2 \left( 1 - \rho^2_{\gamma(\bar{X},\bar{Z})_q} \right) \right]. \quad (12)$$

3. Methodology

3.1. Introducing the estimator schema

Estimator schema, just like database schema in the Software Industry, is a blueprint which serves as guide about the concerned estimator. It is a diagrammatical representation of such estimator. The importance of estimator schema are to ease understanding, abridge any lengthy estimator and to make further modification of concerned estimator easy for samplers.

An instance

$$t^* = \left[ \bar{y}_2 + \sum_{i=1}^{t} \alpha_i (\bar{x}_{(1)i} - \bar{x}_{(2)i}) + \sum_{i=1}^{t} \sigma_i (\bar{x}_i - \bar{x}_{(1)i}) + \sum_{j=t+1}^{k} \gamma_j (\bar{x}_{(1)j} - \bar{x}_{(2)j}) \right]$$

$$+ \sum_{f=k+1}^{g} \beta_f (p_f - p_{(2)f}) \right].$$
The schema for the estimator $t^*$ is presented as

$$t^* = \left[ \begin{array}{c} \overline{y}_2 + \alpha^t_i + \sigma_{1.i}^t + \gamma_j^{k} + \beta_{2.f}^g \\ AV \\ PIC \\ AV \\ PIC \\ AA \\ PIC \end{array} \right]$$

**Line 1**: This explains that the estimator $t^*$ is a partial information case and the type of estimation method involved is regression estimation method.

**Line 2**: $\alpha, \sigma, \gamma$, and $\beta$ are parameters to be estimated.

$\star i, j,$ and $f$ are counters associated with the corresponding parameter.

$i = 1, 2, \ldots, t$, $j = t + 1, \ldots, k$, and $f = k + 1, \ldots, g$.

$\star 1.i$ : FIC with the first phase sample data available.

$\star 2.f$ : FIC with the second phase sample data available.

**Line 3**: This is the type of auxiliary information used. $AV$ means auxiliary variable and $AA$ means auxiliary attribute.

**Line 4**: Explains the type of information case based on the type of auxiliary information being used.

$PIC$ means partial information case, $FIC$ means full information case, and $NIC$ means no information case.

### 3.1.1. Introducing the estimator scheme for full information, no information, and partial information cases

We hereby introduce estimator schema of the aforementioned estimators as proposed by Kung’u et al. [7] as thus:

**a) Estimator schema for full information case (FIC)**

The schema of estimator $t_{MR(3.0)}$ (in Equation (7)) is presented as

$$t^*_{MR(3.0)} = \left[ \begin{array}{c} \overline{y}_2 + \alpha_{2,i}^{k} + \beta_{2,j}^{q} \\ AV \\ (FIC) \\ AA \\ (FIC) \end{array} \right],$$

(14)
where \( AA = \) auxiliary attribute; \( AV = \) auxiliary variable; \( PIC = \) full information case.

(b) Estimator schema for no information case (NIC)

The schema of estimator \( t_{MR(3.2)} \) (in Equation (9)) is presented as

\[
t_{MR(3.2)}^* = \begin{cases} \overline{y}_{2i} + \alpha_i^k + \beta_j^q \\ \overline{y}_{2i} + \alpha_i^k + \beta_j^q \\ \overline{y}_{2i} + \alpha_i^k + \beta_j^q \end{cases} ,
\]

where \( AA = \) auxiliary attribute; \( AV = \) auxiliary variable; \( NIC = \) no information case.

(c) Estimator schema for partial information case (PIC)

The schema of estimator \( t_{MR(3.1)} \) (in Equation (11)) is presented as

\[
t_{MR(3.1)}^* = \begin{cases} \overline{y}_{2i} + \alpha_i^t + \alpha_j^k + \beta_j^q + \gamma_{1f}^q + \beta_m^q \\ \overline{y}_{2i} + \alpha_i^t + \alpha_j^k + \beta_j^q + \gamma_{1f}^q + \beta_m^q \end{cases} ,
\]

where \( AA = \) auxiliary attribute; \( AV = \) auxiliary variable; \( PIC = \) partial information case.

(d) Estimator schema description

\( \overline{y}_{2i} = \) Sample mean of the study variable at the second phase.

\( \sigma_{1,i}^k = \sum_{i=1}^{k} \alpha_{i} (\overline{x}_i - \overline{x}_{(1)i}) \): This is full information case with first phase sample size available.

\( \beta_{2,j}^k = \sum_{j=k+1}^{q} \beta_{j} (\overline{x}_j - \overline{x}_{(2)j}) \): This is full information case with second phase sample data available. Note that \( j \) starts at the end of the value of the last counter. The last counter \( i = 1 \) to \( k \), then \( j = (k + 1) \) to \( q \).

\( \sigma_{i}^k = \sum_{i=1}^{k} \alpha_{i} (\overline{x}_{(1)i} - \overline{x}_{(2)i}) \): This is no information case.
3.2. Proposed mixture regression estimator in two-phase sampling for partial information case II (PIC-II)

If our interest is to estimate the population mean for a mixture regression estimator using multi-auxiliary variables and attributes in two-phase sampling when all the population information on the \( t \) auxiliary variables are available but all the population information on the auxiliary attributes are not available, then we suggest the estimator:

\[
t^2_{OS} = \left\{ \bar{y}_2 + \sum_{i=1}^{k} \alpha_i (\bar{x}_{(1)i} - \bar{x}_{(2)i}) + \sum_{i=1}^{k} \sigma_i (\bar{X}_i - \bar{x}_{(1)i}) + \sum_{j=k+1}^{\ell} \beta_j (p_{(1)j} - p_{(2)j}) 
\right.
\]

\[
+ \sum_{j=k+1}^{\ell} \gamma_j (p_j - p_{(1)j}) + \sum_{m=g+1}^{q} \beta_m (p_{(1)m} - p_{(2)m}) \}.
\]

(17)

The schema for estimator \( t^2_{OS} \) is presented as:

\[
t^2_{OS} = \left\{ \begin{array}{c}
\bar{y}_2 + \alpha_i^{AV} + \sigma_i^{AV} + \beta_j^{AV} + \gamma_j^{AV} + \beta_m^{AV} \\
\end{array} \right\},
\]

(18)

where \( AA \) = auxiliary attribute; \( AV \) = auxiliary variable; \( PIC \) = partial information case; \( FIC \) = full information case.

Applying the Equations (2) and (5) to Equation (17) yields

\[
MSE(t^2_{OS}) = E_1 E_{2/1} \left[ \bar{y}_2 + \sum_{i=1}^{k} \alpha_i (\bar{x}_{(1)i} - \bar{x}_{(2)i}) 
\right.
\]

\[
- \sum_{i=1}^{k} \sigma_i \bar{X}_i + \sum_{j=k+1}^{\ell} \beta_j (\bar{t}_{(1)j} - \bar{t}_{(2)j}) 
\]

\[
- \sum_{j=k+1}^{\ell} \gamma_j \bar{t}_{(1)j} + \sum_{m=g+1}^{q} \beta_m (\bar{t}_{(1)m} - \bar{t}_{(2)m}) \right]^{-2}. \]

(19)
To obtain the optimum values for $\alpha_i$, $\sigma_i$, $\beta_j$, $\gamma_j$, and $\beta_m$, we perform the partial derivative with respect to $\alpha_i$, $\sigma_i$, $\beta_j$, $\gamma_j$, and $\beta_m$ and equate it to zero, hence, solve for the parameters. The optimum equations of the parameters are:

$$\alpha_i = \frac{\bar{Y}_C \gamma (-1)^{i+1} |R_{yx_i}|_{y_{x_k}}}{\bar{X}_i C_{x_i} |A_{x_k}}, \quad \text{for } i = 1, 2, \ldots, k, \quad (20.1)$$

$$\beta_j = \frac{\bar{Y}_C \gamma (-1)^{j+1} |R_{y\tau_j}|_{y_{z_g}}}{P_j C_{\tau_j} |A_{z_g}}, \quad \text{for } j = k+1, k+2, \ldots, g, \quad (20.2)$$

$$\beta_m = \frac{\bar{Y}_C \gamma (-1)^{m+1} |R_{y\tau_m}|_{y_{z_q}}}{P_m C_{\tau_m} |A_{z_q}}, \quad \text{for } m = g+1, g+2, \ldots, q, \quad (20.3)$$

$$\sigma_i = \frac{\bar{Y}_C \gamma (-1)^{i+1} |R_{yx_i}|_{y_{x_k}}}{\bar{X}_i C_{x_i} |A_{x_k}}, \quad \text{for } i = 1, 2, \ldots, k, \quad (20.4)$$

$$\gamma_j = \frac{\bar{Y}_C \gamma (-1)^{j+1} |R_{y\tau_j}|_{y_{z_g}}}{P_j C_{\tau_j} |A_{z_g}}, \quad \text{for } j = k+1, k+2, \ldots, g. \quad (20.5)$$

Applying expectation, hence, insert the optimum equations obtained for $\alpha_i$, $\sigma_i$, $\beta_j$, $\gamma_j$, and $\beta_m$

$$MSE (\hat{\theta}_{OS}) = \bar{Y}^2 C_y \left[ \theta_2 \left( 1 - \rho^2_{y(\bar{X}, z)_q} \right) + \theta_1 \left( \rho^2_{y(\bar{X}, z)_q} - \rho^2_{y(\bar{X}, z)_g} \right) \right]. \quad (21)$$
3.3. Proposed mixture regression estimator in two-phase sampling for partial information case III (PIC-III)

If our interest is to estimate the population mean for a mixture regression estimator using multi-auxiliary variables and attributes in two-phase sampling when all the population information on the auxiliary variables are not available but all the population information on the auxiliary attributes are available, then we suggest the estimator:

\[
\begin{align*}
t_{OS}^3 &= \left\{ \bar{y}_2 + \sum_{i=1}^{t} \alpha_i \left( \bar{x}_{(1)} - \bar{x}_{(2)} \right) + \sum_{i=1}^{t} \sigma_i \left( \bar{x}_i - \bar{x}_{(1)} \right) + \sum_{j=t+1}^{k} \alpha_j \left( \bar{x}_{(1)} - \bar{x}_{(2)} \right) \right. \\
&\quad + \sum_{j=k+1}^{q} \beta_f \left( p_{(1)} - p_{(2)} \right) + \sum_{f=k+1}^{q} \gamma_f \left( p_f - p_{(1)} \right) \right\}. \tag{22}
\end{align*}
\]

The schema for estimator \( t_{OS}^3 \) is presented as:

\[
\begin{align*}
t_{OS}^{3*} &= \left\{ \bar{y}_2 + \alpha_i + \sigma_i + \beta_f + \gamma_f \right\}, \tag{23}
\end{align*}
\]

where \( AA = \) auxiliary attribute; \( AV = \) auxiliary variable; \( PIC = \) partial information case; \( FIC = \) full information case.

Applying the Equations (2) and (5) to Equation (22) gives

\[
\begin{align*}
\text{MSE}(t_{OS}^3) &= E_1 E_{2/1} \left[ \bar{y}_2 + \sum_{i=1}^{t} \alpha_i \left( \bar{\bar{x}}_{x(1)} - \bar{\bar{x}}_{x(2)} \right) \right. \\
&\quad - \sum_{i=1}^{t} \sigma_i \bar{\bar{x}}_{x(1)} + \sum_{j=t+1}^{k} \alpha_j \left( \bar{\bar{x}}_{x(1)} - \bar{\bar{x}}_{x(2)} \right) \right. \\
&\quad + \sum_{j=k+1}^{q} \beta_f \left( \bar{\bar{e}}_{x(1)} - \bar{\bar{e}}_{x(2)} \right) - \sum_{f=k+1}^{q} \gamma_f \bar{\bar{e}}_{x(1)} \left( \bar{\bar{e}}_{x(2)} \right) \right\}^2. \tag{24}
\end{align*}
\]
To obtain the optimum values for $\alpha_i$, $\sigma_i$, $\alpha_j$, $\beta_f$, and $\gamma_f$, we perform the partial derivative with respect to $\alpha_i$, $\sigma_i$, $\alpha_j$, $\beta_f$, and $\gamma_f$ and equate it to zero. The optimum equations for the parameters are:

$$
\alpha_i = \frac{\bar{y}_{C_y} (-1)^{i+1} \left| R_{y,x_i} \right|_{y=\bar{y}_x}}{\bar{x}_i \bar{C}_{x_i} \left| R_{y,y} \right|_{y=\bar{y}_x}}, \quad \text{for } i = 1, 2, \ldots, t,
$$

(25.1)

$$
\alpha_j = \frac{\bar{y}_{C_y} (-1)^{j+1} \left| R_{y,x_j} \right|_{y=\bar{y}_x}}{\bar{x}_j \bar{C}_{x_j} \left| R_{y,y} \right|_{y=\bar{y}_x}}, \quad \text{for } j = t + 1, t + 2, \ldots, k,
$$

(25.2)

$$
\beta_f = \frac{\bar{y}_{C_y} (-1)^{f+1} \left| R_{y,\tau_f} \right|_{y=\bar{y}_x}}{P_f \bar{C}_{\tau_f} \left| R_{y,y} \right|_{y=\bar{y}_x}}, \quad \text{for } f = k + 1, k + 2, \ldots, q,
$$

(25.3)

$$
\sigma_i = \frac{\bar{y}_{C_y} (-1)^{i+1} \left| R_{y,x_i} \right|_{y=\bar{y}_x}}{\bar{x}_i \bar{C}_{x_i} \left| R_{y,y} \right|_{y=\bar{y}_x}}, \quad \text{for } i = 1, 2, \ldots, t,
$$

(25.4)

$$
\gamma_f = \frac{\bar{y}_{C_y} (-1)^{f+1} \left| R_{y,\tau_f} \right|_{y=\bar{y}_x}}{P_f \bar{C}_{\tau_f} \left| R_{y,y} \right|_{y=\bar{y}_x}}, \quad \text{for } f = k + 1, k + 2, \ldots, q.
$$

(25.5)

Simplify Equation (24), and insert the optimum equations obtained for $\alpha_i$, $\sigma_i$, $\alpha_j$, $\beta_f$, and $\gamma_f$

$$
MSE \left( \theta^{\text{OS}}_3 \right) = \bar{y}_{C_y}^2 \cdot \theta_2 \left( 1 - \rho_{y.(\bar{y},\bar{z})}^2 \right) + \theta_1 \left( \rho_{y.(\bar{y},\bar{z})}^2 - \rho_{y.(\bar{y},\bar{z})}^{2*} \right).
$$

(26)
4. Results and Discussion

4.1. Theoretical comparison of the estimators in PIC: case-I, case-II and case-III

4.1.1. Comparison of PIC-I and PIC-II

\[
MSE(t_{MR(3,1)}) - MSE(t_{OS}^2),
\]

\[
\theta_2 \rho_{yx_t}^2 - \theta_1 \rho_{yx_h}^2 > 0.
\]

Equation (28) remains to be true since \((\theta_2 > \theta_1)\), and \(0 \leq \rho_{yx_t}^2 \leq 1\) and \(0 \leq \rho_{yx_h}^2 \leq 1\), though \(\rho_{yx_t}^2 > \rho_{yx_h}^2\) is expected because \(t\) auxiliary variables is a subset of \(k\) auxiliary variables, hence, justifying for the inclusion of new auxiliary variables. Empirical analysis based on simulated data in Table 1 confirms the validity of Equation (28). This implies that \(MSE(t_{MR(3,1)}) > MSE(t_{OS}^2)\). Hence, our proposed estimator, \(t_{OS}^2\), is efficient over \(t_{MR(3,1)}\) estimator.
Table 1. Empirical comparison of PIC-I and PIC-II

| SN | $n_1$ | $n_2$ | $\theta_1 = \frac{1}{n_1}$ | $\theta_2 = \frac{1}{n_2}$ | $\rho_{y,x_1}^2$ | $\rho_{y,x_2}^2$ | $\theta_2 \rho_{y,x_j}^2 - \theta_1 \rho_{y,x_h}^2$ | Decision |
|----|------|------|----------------|----------------|----------------|----------------|-----------------|----------|
| 1  | 10000| 2000 | 0.00010        | 0.00050        | 0.5099         | 0.5604         | 0.000199        | $\text{MSE}(t_{MR(3.1)}) > \text{MSE}(t_{OS}^2)$ |
| 2  | 5000 | 1000 | 0.00020        | 0.00100        | 0.5021         | 0.6870         | 0.000365        | $\text{MSE}(t_{MR(3.1)}) > \text{MSE}(t_{OS}^2)$ |
| 3  | 2000 | 400  | 0.00050        | 0.00250        | 0.7666         | 0.8678         | 0.001483        | $\text{MSE}(t_{MR(3.1)}) > \text{MSE}(t_{OS}^2)$ |
| 4  | 1000 | 200  | 0.00100        | 0.00500        | 0.5233         | 0.5610         | 0.002056        | $\text{MSE}(t_{MR(3.1)}) > \text{MSE}(t_{OS}^2)$ |
| 5  | 800  | 160  | 0.00125        | 0.00625        | 0.5179         | 0.5886         | 0.002501        | $\text{MSE}(t_{MR(3.1)}) > \text{MSE}(t_{OS}^2)$ |
| 6  | 450  | 90   | 0.00222        | 0.01111        | 0.6088         | 0.6333         | 0.005357        | $\text{MSE}(t_{MR(3.1)}) > \text{MSE}(t_{OS}^2)$ |
| 7  | 400  | 80   | 0.00250        | 0.01250        | 0.6961         | 0.7980         | 0.006707        | $\text{MSE}(t_{MR(3.1)}) > \text{MSE}(t_{OS}^2)$ |
| 8  | 350  | 70   | 0.00286        | 0.01429        | 0.5328         | 0.6512         | 0.005751        | $\text{MSE}(t_{MR(3.1)}) > \text{MSE}(t_{OS}^2)$ |
| 9  | 300  | 60   | 0.00333        | 0.01667        | 0.5327         | 0.6581         | 0.006684        | $\text{MSE}(t_{MR(3.1)}) > \text{MSE}(t_{OS}^2)$ |
| 10 | 250  | 50   | 0.00400        | 0.02000        | 0.5198         | 0.5581         | 0.008165        | $\text{MSE}(t_{MR(3.1)}) > \text{MSE}(t_{OS}^2)$ |
4.1.2. Comparison of PIC-I and PIC-III

\[ MSE(t_{MR(3.1)}) - MSE(t_{OS}^3), \]  

\[ \theta_2 \rho_{y^2 q}^2 - \theta_1 \rho_{y^2 q}^2* > 0. \]  

Equation (30) remains to be true since \((\theta_2 > \theta_1)\), and \(0 \leq \rho_{y^2 q}^2 \leq 1\) and \(0 \leq \rho_{y^2 q}^2* \leq 1\), though \(\left\{ \rho_{y^2 q}^2* > \rho_{y^2 q}^2 \right\}\) is expected because \(g\) auxiliary attributes is a subset of \(q\) auxiliary attributes, hence, justifying for the inclusion of new auxiliary attributes. Empirical analysis based on simulated data in Table 2 confirms the validity of Equation (30). This implies that \(MSE(t_{MR(3.1)}) > MSE(t_{OS}^3)\). Hence, our proposed estimator, \(t_{OS}^3\), is efficient over \(t_{MR(3.1)}\) estimator.
Table 2. Empirical comparison of PIC-I and PIC-III

| SN | $n_1$ | $n_2$ | $\theta_1 = \frac{1}{n_1}$ | $\theta_2 = \frac{1}{n_2}$ | $\rho_{\gamma,y}^2$ | $\rho_{\gamma,y}^{2*}$ | $\theta_2 \rho_{\gamma,y}^{2*} - \theta_1 \rho_{y,g}^{2*}$ | Decision |
|----|------|------|-----------------|-----------------|-------------|-------------|------------------------|---------|
| 1  | 10000| 2000 | 0.00010         | 0.00050         | 0.5178      | 0.6702      | 0.000192               | $MSE(\mu_{(3,1)}) > MSE(\mu_{OS})$ |
| 2  | 5000 | 1000 | 0.00020         | 0.00100         | 0.5802      | 0.6900      | 0.000442               | $MSE(\mu_{(3,1)}) > MSE(\mu_{OS})$ |
| 3  | 2000 | 400  | 0.00050         | 0.00250         | 0.5635      | 0.5853      | 0.001116               | $MSE(\mu_{(3,1)}) > MSE(\mu_{OS})$ |
| 4  | 1000 | 200  | 0.00100         | 0.00500         | 0.5340      | 0.7131      | 0.001957               | $MSE(\mu_{(3,1)}) > MSE(\mu_{OS})$ |
| 5  | 800  | 160  | 0.00125         | 0.00625         | 0.7413      | 0.7849      | 0.003652               | $MSE(\mu_{(3,1)}) > MSE(\mu_{OS})$ |
| 6  | 450  | 90   | 0.00222         | 0.01111         | 0.5199      | 0.5474      | 0.004560               | $MSE(\mu_{(3,1)}) > MSE(\mu_{OS})$ |
| 7  | 400  | 80   | 0.00250         | 0.01250         | 0.6348      | 0.6653      | 0.006272               | $MSE(\mu_{(3,1)}) > MSE(\mu_{OS})$ |
| 8  | 350  | 70   | 0.00286         | 0.01429         | 0.8534      | 0.8926      | 0.009641               | $MSE(\mu_{(3,1)}) > MSE(\mu_{OS})$ |
| 9  | 300  | 60   | 0.00333         | 0.01667         | 0.6973      | 0.8472      | 0.008798               | $MSE(\mu_{(3,1)}) > MSE(\mu_{OS})$ |
| 10 | 250  | 50   | 0.00400         | 0.02000         | 0.7018      | 0.7531      | 0.011023               | $MSE(\mu_{(3,1)}) > MSE(\mu_{OS})$ |
4.1.3. Comparison of PIC-II and PIC-III

\[ t_{OS}^3 \] would be efficient over \( t_{OS}^2 \) if \( MSE(t_{OS}^2) - MSE(t_{OS}^3) \). \hspace{1cm} (31)

\[
\begin{pmatrix}
\theta_1 \rho_{y,x_k}^{2*} - \theta_2 \rho_{y,x_l}^{2*} \\
\theta_2 \rho_{y,x_l}^{2*} - \theta_1 \rho_{y,x_k}^{2*}
\end{pmatrix}.
\hspace{1cm} (32)
\]

where \( \theta_2 > \theta_1 \), \( 0 \leq \rho_{y,x_k}^{2*} \leq 1 \), \( 0 \leq \rho_{y,x_l}^{2*} \leq 1 \), \( 0 \leq \rho_{y,x_q}^{2*} \leq 1 \) and \( 0 \leq \rho_{y,x_g}^{2*} \leq 1 \).

From Equation (32), it is expected that \( \rho_{y,x_k}^{2*} > \rho_{y,x_l}^{2*} \) and \( \rho_{y,x_q}^{2*} > \rho_{y,x_g}^{2*} \).

Consequently, it is expected that Equation (32) to be less than zero. If this holds then

\[ MSE(t_{OS}^2) < MSE(t_{OS}^3). \hspace{1cm} (33) \]

This implies that estimator \( t_{OS}^2 \) is efficient over \( t_{OS}^3 \). Empirical analysis based on simulated data in Table 3 confirms Equation (33) to be true. Hence, PIC-II estimator is efficient over PIC-III estimator.
Table 3. Empirical comparison of PIC-II and PIC-III

| SN | $n_1$ | $n_2$ | $\theta_1 = \frac{1}{n_1}$ | $\theta_2 = \frac{1}{n_2}$ | $\rho^2_{y_1 \xi_1}$ | $\rho^2_{y_2 \xi_2}$ | $\rho^2_{y_3 \xi_3}$ | $\rho^2_{y_4 \xi_4}$ | $\left( \frac{\theta_1 \rho^2_{y_5 \xi_5} - \theta_2 \rho^2_{y_6 \xi_6}}{\rho^2_{y_7 \xi_7} - \theta_2 \rho^2_{y_8 \xi_8}} \right)$ | Decision |
|----|-------|-------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|--------------------------------|-----------|
| 1  | 10000 | 2000  | 0.00010         | 0.00050         | 0.5072          | 0.5149          | 0.5382          | 0.5735          | -0.000414                  | $t^2_{OS}$ is efficient |
| 2  | 5000  | 1000  | 0.00020         | 0.00100         | 0.5276          | 0.5991          | 0.5554          | 0.6136          | -0.000840                  | $t^2_{OS}$ is efficient |
| 3  | 2000  | 400   | 0.00050         | 0.00250         | 0.5315          | 0.5991          | 0.8325          | 0.8936          | -0.002664                  | $t^2_{OS}$ is efficient |
| 4  | 1000  | 200   | 0.00100         | 0.00500         | 0.5031          | 0.5673          | 0.5599          | 0.6038          | -0.004144                  | $t^2_{OS}$ is efficient |
| 5  | 800   | 160   | 0.00125         | 0.00625         | 0.8654          | 0.8742          | 0.5004          | 0.6476          | -0.006634                  | $t^2_{OS}$ is efficient |
| 6  | 450   | 90    | 0.00222         | 0.01111         | 0.5856          | 0.7884          | 0.5066          | 0.5375          | -0.009189                  | $t^2_{OS}$ is efficient |
| 7  | 400   | 80    | 0.00250         | 0.01250         | 0.5178          | 0.5530          | 0.5182          | 0.7518          | -0.009688                  | $t^2_{OS}$ is efficient |
| 8  | 350   | 70    | 0.00286         | 0.01429         | 0.7008          | 0.7941          | 0.5659          | 0.8885          | -0.013288                  | $t^2_{OS}$ is efficient |
| 9  | 300   | 60    | 0.00333         | 0.01667         | 0.5378          | 0.5411          | 0.5182          | 0.5736          | -0.013884                  | $t^2_{OS}$ is efficient |
| 10 | 250   | 50    | 0.00400         | 0.02000         | 0.6609          | 0.8308          | 0.5392          | 0.5504          | -0.018479                  | $t^2_{OS}$ is efficient |
5. Conclusion

Our two proposed partial information cases (PIC-II and PIC-III) estimators gain improved efficiency over the partial information case I (PIC-I) as established by Kung’u et al. [7]. However, PIC-II is efficient over PIC-III. This means that, it is better to have an PIC estimator with combination of FIC of auxiliary variables and PIC of auxiliary attributes as against PIC estimator with combination of PIC of auxiliary variables combined with FIC of auxiliary attributes.

References

[1] Ahmad Zahoor, Shabbaz Muhammad Qaiser and Hanif Muhammad, Two-Phase Sampling, Cambridge Scholars Publishing, First Edition, ISBN (10): 1-4438-4595-7. ISBN(13) (2013), 978-1-4438-4595-3.

[2] S. Arora and L. Bans, New Mathematical Statistics, Satya Prahashar, New Dehli, 1989.

[3] S. Bahl and R. K. Tuteja, Ratio and product type exponential estimators, Journal of Information and Optimization Science 12(1) (1991), 159-164.

http://dx.doi.org/10.1080/02522667.1991.10699058

[4] A. K. Das, Contribution to the Theory of Sampling Strategies based on Auxiliary Information, Ph.D. Thesis submitted to B. C. K. V. Mohanpur, Nadia, West Bengal, India, 1988.

[5] M. Hanif, U. HaqI and M. Q. Shahbaz, On a new family estimators using multiple auxiliary attributes, World Applied Sciences Journal 7(11) (2009), 1419-1422.

[6] H. S. Jhajj, M. K. Sharma and L. K. Grover, A family of estimator of population mean using information on auxiliary attributes, Pakistan Journal of Statistics 22 (2006), 43-50.

[7] J. Kung’u, G. Chumba and L. Odongo, Mixture regression estimators using multi-auxiliary variables and attributes in two-phase sampling, Open Journal of Statistics 4 (2014), 355-366.

http://dx.doi.org/10.4236/ojs.2014.45035

[8] M. Moeen, Q. Shahbaz and M. Hanif, Mixture ratio and regression estimators using multi-auxiliary variables and attributes in single phase sampling, World Applied Sciences Journal 18 (2012), 1518-1526.

[9] J. Neyman, On the two aspects of representing method of stratified sampling and the method of purposive selection, J. Roy. Statist. Soc. 97 (1934), 558-606.
[10] J. Neyman, Contribution of the theory of sampling human populations, J. Amer. Statist. Assoc. 33 (1938), 101-116.

[11] I. Olkin, Multivariate ratio estimation for finite population, Biometrika VI 45 (1958), 154-165.

http://dx.doi.org/10.1093/biomet/45.1-2.154

[12] S. Rajesh, C. Pankaj, S. Nirmala and S. Florentins, Ratio-Product Type Exponential Estimator or Estimating Finite Population Mean using Information on Auxiliary attributes, Renaissance, High Press, USA, 2007.

[13] Samiuddin and Hanif, Estimation of population mean in single phase sampling with or without additional information, Pakistan Journal of Statistics 23 (2007), 99-118.

[14] T. P. Tripathi, Contributions of the Sampling Theory using Multivariate Information, Ph.D. Thesis submitted to Punjabi University, Patiala, India, 1970.

[15] D. J. Watson, The estimation of leaf area, J. Agri. Sci. 27 (1937), 475.