Transactions of the Nigerian Association of Mathematical Physics
Vol. 6 **(Jan., 2018), pp. 179– 189**

In honour of Prof. Ekhaguere at 70

Extension of information cases on mixture ratio estimators using multi-auxiliary variables and attributes in two-phase sampling

P. I. Ogunyinka[a*] **and A. A. Sodipo**[b]

[a]*Department of Mathematical Sciences, Olabisi Onabanjo University, Ago-Iwoye, Nigeria*;
[b]*Department of Statistics, University of Ibadan, Ibadan, Nigeria*

Abstract. The use of multiple auxiliary characters like auxiliary attributes and variables has been confirmed to improve the efficiency of estimators in two-phase sampling. The three ways of utilizing auxiliary characters (full, partial and no information cases) have provided flexibility in its usage of these auxiliary characters. In this article, we have proposed two additional partial information cases (PIC-II and PIC-III) to the existing partial information case (PIC-I) of using multi-auxiliary attributes and variables in two-phase sampling. The estimator schemas were introduced to the five estimators for the purpose of estimator formation. It was ascertained that our two proposed estimators PIC-II and PIC-III were efficient over the existing partial information case estimators (PIC-I) subject to the conditions of availability of the number of auxiliary variables and auxiliary attributes in the PIC-I estimator. The condition under which PIC-II is equally efficient as PIC-III was established. However, we subject the use of PIC-II or PIC-III to the availability of auxiliary variables and auxiliary attributes. Finally, our proposed estimators proved efficient over No Information Case (NIC) estimator in our review.

Keywords: partial information case, mixture ratio estimator, two-phase sampling, auxiliary attributes, auxiliary variables.

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 (1934, 1938) initiated the use of auxiliary variable at the pre-selection stage while Cochran (1940) first coined ratio estimator at the post-selection stage. The use of ratio estimator in two-phase sampling towards estimating the study variable uses the auxiliary information at the post-selection or estimation stage. Olkin (1958) 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.* (2013) summarized the works of Tripathi (1970) and Das (1988) into four ways which auxiliary information may be available in two-phase sampling (availability of auxiliary information). Among these four ways are when exact values of the parameters are not known but their estimated values are known (called No Information Case) and the values of one or more parameters of auxiliary variables may be known (called Full Information Case). Samiuddin and Hanif (2007) introduced the third information case called Partial Information Case (PIC) which is the combination of both full information and no information cases into ratio and regression estimation methods. 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 character about the population could be dichotomous property (present or absent) which also highly correlated with the study variable. The use of such dichotomous character, called auxiliary attribute, in Sample Survey has revealed improvement on the estimation of the study variable. Bahl and Tuteja (1991), Jhajj *et al.* (2006), Rajesh *et al.* (2007), Hanif *et al.* (2009) and

*Corresponding author. Email: ogunyinka.peter@gmail.com

Moeen *et al.* (2012) are among the literatures that have utilized auxiliary attributes to obtain improved estimators over the singular use of auxiliary variable and over non-usage of auxiliary variable. Mixture estimator uses the combination of auxiliary variables and attributes towards the improvement of an estimator. Waweru *et al.* (2014) proposed generalized mixture Ratio estimators in two-phase sampling with the combination of multiple auxiliary variables and attributes following the prior three ways of the availability of the auxiliary variables (Full, No and Partial Information Cases) as established by Samiuddin and Hanif (2007). Waweru *et al.* (2014) estimators gained efficiency over the prior reviewed estimators of Hanif *et al.* (2009).

Considering the partial information case established by Waweru et al. (2014), this article considers two additional cases of such partial information cases to make up three partial information cases (PIC-I, PIC-II and PIC-III). Consequently, there are five information cases subject to the inclusion of full and no information cases. The estimator schema for the five information cases were introduced and the corresponding mean square errors were established following Arora and Bansi (1989) approach of presenting mean square error.

2. Preliminaries

2.1 *Notation and assumption*

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 for where $n_1 > n_2$. Hence, presenting

$$\theta_1 = \left(\frac{1}{n_1} - \frac{1}{N}\right); \theta_2 = \left(\frac{1}{n_2} - \frac{1}{N}\right); for\left(\theta_1 < \theta_2\right) \tag{2.1}$$

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 sample. Then

$$\overline{y}_2 = \left(\overline{Y} + \overline{e}_{y2}\right); \overline{x}_{(1)i} = \left(\overline{X}_i + \overline{e}_{x(1)i}\right); \overline{x}_{(2)i} = \left(\overline{X}_i + \overline{e}_{x(2)i}\right); \text{ for } i = 1, 2, \cdots, p \tag{2.2}$$

where $\overline{e}_{y2}$, $\overline{e}_{x(1)i}$ and $\overline{e}_{x(2)i}$ are the mean sampling errors and are very small, such that

$$E\left(\overline{e}_{x(1)i}\right) = E\left(\overline{e}_{x(2)i}\right) = 0; \tag{2.3}$$

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

$$f(x) = \begin{cases} 1 & j^{th} \text{ unit of population possessing } i^{th} \text{ auxiliary attributes} \\ 0 & \text{Otherwise} \end{cases} \tag{2.4}$$

τ_j = value of j_{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 τj. Let $P_j = \frac{A_j}{N}$ and $p_j = \frac{a_j}{n}$ be the corresponding population and sample proportion possessing attribute τj. Similarly,

$$p_{(1)i} = \left(P_i + \overline{e}_{\tau(1)i}\right); \; p_{(2)i} = \left(P_i + \overline{e}_{\tau(2)i}\right); \tag{2.5}$$

$$\text{for } E\left(\overline{e}_{\tau(1)i}\right) = E\left(\overline{e}_{\tau(2)i}\right) = 0; \tag{2.6}$$

and $C_y^2 = \frac{S_y^2}{\overline{Y}^2}$; $C_{\tau 1}^2 = \frac{S_{\tau 1}{}^2}{P^2}$; $\rho_{yx} = \frac{S_{yx}}{S_y S_x}$

2.2 *Some other useful results*

Similarly, the following results are also necessary in establishing the mean square errors of our proposed estimators.

$E\left(\overline{e}_{y2}\right)^2 = \theta_2\overline{Y}^2C_y^2$; $E\left(\overline{e}_{x(2)i}\right)^2 = \theta_2\overline{X}_i^2C_{xi}^2$; $E\left(\overline{e}_{y2}\overline{e}_{x(2)i}\right) = \theta_2\overline{YX}_iC_yC_{xi}\rho_{yxi}$

$E\left(\overline{e}_{y2}\left(\overline{e}_{x(1)i} - \overline{e}_{x(2)i}\right)\right) = \left(\theta_1 - \theta_2\right)\overline{YX}_iC_yC_{xi}\rho_{yxi}$

$E\left(\overline{e}_{x(2)i}\left(\overline{e}_{x(1)i} - \overline{e}_{x(2)i}\right)\right) = \left(\theta_1 - \theta_2\right)\overline{X}_i^2C_{xi}^2$

$E\left(\overline{e}_{x(1)i}\left(\overline{e}_{x(1)i} - \overline{e}_{x(2)i}\right)\right) = 0$

$E\left(\overline{e}_{x(1)i} - \overline{e}_{x(2)i}\right)^2 = \left(\theta_2 - \theta_1\right)\overline{X}_i^2C_{xi}^2$

$E\left(\left(\overline{e}_{x(1)i} - \overline{e}_{x(2)i}\right)\left(\overline{e}_{x(1)j} - \overline{e}_{x(2)j}\right)\right) = \left(\theta_2 - \theta_1\right)\overline{X}_i\overline{X}_jC_{xi}C_{xj}\rho_{xixj}$ for $i \neq j$

$E\left(\overline{e}_{x(1)i}\overline{e}_{x(1)j}\right) = \theta_1\overline{X}_i\overline{X}_jC_{x_i}C_{x_j}\rho_{x_ix_j}$ for $i \neq j$

$E\left(\overline{e}_{x(1)i}\overline{e}_{x(2)i}\right) = \theta_1\overline{X}_i\overline{X}_jC_{x_i}C_{x_j}\rho_{x_ix_j}$ for $i \neq j$

$E\left(\overline{e}_{y_2}\overline{e}_{x(1)i}\right) = \theta_1\overline{YX}_iC_yC_{x_i}\rho_{yx_i}$

$E\left(\overline{e}_{\tau(1)i} - \overline{e}_{\tau(2)i}\right)^2 = \left(\theta_2 - \theta_1\right)P_i^2C_{\tau_i}^2$

$E\left(\overline{e}_{y_2}\overline{e}_{\tau(2)i}\right) = \theta_2\overline{Y}P_iC_yC_{\tau_i}\rho_{y\tau_i}$

$E\left(\overline{e}_{\tau(2)_i}\left(\overline{e}_{\tau(1)_i} - \overline{e}_{\tau(2)i}\right)\right) = \left(\theta_1 - \theta_2\right)P_i^2C_{\tau_i}^2$

$E\left(\overline{e}_{\tau(2)i}\overline{e}_{\tau(2)j}\right) = \theta_2P_iP_jC_{\tau_i}C_{\tau_j}\rho_{\tau_i\tau_j}$ for $i \neq j$

$E\left(\left(\overline{e}_{\tau(1)i} - \overline{e}_{\tau(2)i}\right)\left(\overline{e}_{\tau(1)_j} - \overline{e}_{\tau(2)j}\right)\right) = \left(\theta_2 - \theta_1\right)P_iP_jC_{\tau_i}C_{\tau_j}\rho_{\tau_i\tau_j}$ for $i \neq j$

$E\left(\overline{e}_{\tau(2)i}\left(\overline{e}_{\tau(1)j} - \overline{e}_{\tau(2)j}\right)\right) = \left(\theta_1 - \theta_2\right)P_iP_jC_{\tau i}C_{\tau j}\rho_{\tau_i}\tau_j$ for $i \neq j$

According to Arora and Bansi (1989)

$$\left(1 - \left[\frac{\sum_{i=1}^{q}\left(-1\right)^{i+1}\left|R_{yxi}\right|_{y\underline{x}_q}\rho_{yxi}}{\left|R\right|_{\underline{x}_q}}\right]\right) = \frac{\left|R\right|_{y\underline{x}_q}}{\left|R\right|_{\underline{x}_q}} = \left(1 - \rho_{y.\underline{x}_q}^2\right)$$

2.3 *Mixture ratio estimator in two-phase sampling*

2.3.1 Full Information Case (FIC)

Waweru *et al.* (2014) established the estimated population mean of a generalized mixture Ratio 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)* and presented as

$$t_1 = \overline{y}_2\prod_{i=1}^{k}\left(\frac{\bar{X}_i}{\bar{x}_{(2)i}}\right)^{\alpha_i}\prod_{j=k+1}^{q}\left(\frac{P_j}{p_{(2)_j}}\right)^{\beta_j} \tag{2.7}$$

The corresponding Mean Square Error (MSE) is presented as

$$MSE\left(t_1\right) = \theta_2\bar{Y}^2C_y^2\left(1 - \rho_{y.\underline{x}_k}^2 - \rho_{y.\underline{\tau}_q}^2\right)$$

$$MSE\left(t_1\right) = \theta_2\bar{Y}^2C_y^2\left(1 - \rho_{y(\underline{x},\underline{\tau})_q}^2\right) \tag{2.8}$$

2.3.2 No Information Case (NIC)

When the population information of all the auxiliary variables and attributes are not available, hence, the estimated population mean of the mixture ratio estimator using multi-auxiliary variables and attributes in two-phase sampling is presented by Waweru *et al.* (2014) as

$$t_2 = \bar{y}_2 \prod_{i=1}^{k} \left(\frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}} \right)^{\alpha_i} \prod_{j=k+1}^{q} \left(\frac{p_{(1)j}}{p_{(2)_j}} \right)^{\beta_j} \tag{2.9}$$

The corresponding Mean Square Error (MSE) is given as

$$MSE\big(t_2\big) = \bar{Y}^2 C_y^2 \Big(\theta_2 + (\theta_1 - \theta_2)\rho^2_{y.\underline{x}_k} + (\theta_1 - \theta_2)\rho^2_{y.\tau_q} \Big)$$

$$MSE\big(t_2\big) = \bar{Y}^2 C_y^2 \left(\theta_2 \left(1 - \rho^2_{y(\underline{x},\underline{\tau})_q} \right) + \theta_2 \rho^2_{y(\underline{x},\underline{\tau})_q} \right) \tag{2.10}$$

2.3.3 Partial Information Case (PIC)

Waweru *et al.* (2014) presented the estimated population mean of mixture ratio 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. Waweru *et al.* (2014) used the second method of configuring partial information case out of the two ways expressed by Ahmed *et al.* (2013) of presenting partial Information case. The estimator is presented as:

$$t_3 = \bar{y}_2 \left[\prod_{i=1}^{r} \left(\frac{\bar{x}_{(1)_i}}{\bar{x}_{(2)_i}} \right)^{\alpha_i} \left(\frac{\bar{X}_i}{\bar{x}_{(1)_i}} \right)^{\beta_i} \right] \left[\prod_{j=r+1}^{k} \left(\frac{\bar{x}_{(1)_j}}{\bar{x}_{(2)_j}} \right)^{\alpha_j} \right] \left[\prod_{f=k+1}^{h} \left(\frac{p_{(1)_f}}{p_{(2)_f}} \right)^{\gamma_f} \left(\frac{P_f}{p_{(1)_f}} \right)^{\lambda_f} \right] \left[\prod_{g=h+1}^{q} \left(\frac{p_{(1)_g}}{p_{(2)_g}} \right)^{\gamma_g} \right] \tag{2.11}$$

The corresponding Mean Square Error *(MSE)* is given as:

$$MSE\big(t_3\big) = \bar{Y}^2 C_y^2 \Big[\theta_2 - \theta_2 \rho^2_{y.\underline{x}_r} - \theta_2 \rho^2_{y.\underline{x}_k} - \theta_2 \rho^2_{y.\underline{\tau}_h} - \theta_2 \rho^2_{y.\underline{\tau}_q} + \theta_1 \rho^2_{y.\underline{x}_k} + \theta_2 \rho^2_{y.\underline{\tau}_q} \Big] \tag{2.12}$$

This is further simplified as thus:

$$MSE\big(t_3\big) = \bar{Y}^2 C_y^2 \Big[\theta_2 \Big(1 - \rho^2_{y.(\underline{x},\underline{\tau})_q} \Big) + \theta_1 \Big(\rho^2_{y.\underline{x}_k} + \rho^2_{y.\underline{\tau}_q} \Big) \Big] \tag{2.13}$$

3. Methodology

3.1 *Introducing the estimator scheme*

Estimator schema, just like database schema in the Software Industry, is a blue-print which serves as guide about the concerned estimator. It is a diagrammatic 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. Example of a ratio estimator in two-phase sampling is:

$$t = \bar{y}_2 \left[\prod_{i=1}^{r} \left(\frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}} \right)^{\alpha_i} \left(\frac{\bar{X}}{\bar{x}_{(1)i}} \right)^{\beta_i} \right] \left[\prod_{j=r+1}^{q} \left(\frac{P_j}{x_{(2)_j}} \right)^{\lambda_j} \right]$$

The Schema for the estimator t is presented as

$$t^* = \left[\overbrace{\bar{y}_2 * \underbrace{\alpha_i^{+r} * \beta_{1.i}^{+r}}_{AV(PIC)} * \underbrace{\lambda_{2.j}^{+q}}_{AA(FIC)\rightarrow LINE3}}^{Ratio(PIC)\longrightarrow LINE1} \longrightarrow LINE2\right]$$

LINE 1: This explains that the estimator t^* is a partial information case and the type of estimation method involved is ratio estimation method.
LINE 2: * α, β and λ are parameters to be estimated in the estimator. * i and j are counters associated with the corresponding parameter. $i = 1, 2, \cdots, r$, $j = r+1, r+2, \cdots, q$
* $1.i$: FIC with the first phase sample data available
* $2.j$: FIC with the second phase sample data available
LINE 3: This is the type of auxiliary information used. **AV** means Auxiliary Variable an **AA** means Auxiliary Attribute. It further 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 estimator schema for full information, no information and partial information cases

We hereby introduce estimator schema of the aforementioned estimators as proposed by Waweru *et al.* (2014) as thus:

(1) Estimator Schema for Full Information Case (FIC). The schema of estimator t_1 is presented as

$$t_1^* = \left[\overbrace{\bar{y}_2 * \underbrace{\alpha_{2.i}^{+k}}_{AV(FIC)} * \underbrace{\beta_{2.j}^{+q}}_{AA(FIC)}}^{Ratio(FIC)}\right] \tag{3.1}$$

(2) Estimator Schema for No Information Case (NIC). The schema of estimator t_2 is presented as

$$t_2^* = \left[\overbrace{\bar{y}_2 * \underbrace{\alpha_i^{+k}}_{AV(NIC)} * \underbrace{\beta_j^{+q}}_{AA(NIC)}}^{Ratio(NIC)}\right] \tag{3.2}$$

(3) Estimator Schema for Partial Information Case (PIC). The schema of estimator t_3 is presented as

$$t_3^* = \left[\overbrace{\bar{y}_2 * \underbrace{\alpha_i^{+r} * \beta_{1.i}^{+r} * \alpha_j^{+k}}_{AV(PIC)} * \underbrace{\gamma_f^{+h} * \lambda_{1.f}^{+h} * \gamma_g^{+q}}_{AA(PIC)}}^{Ratio(PIC-I)}\right] \tag{3.3}$$

(4) Estimator Schema Description. $\bar{y}_2$= Sample mean of the study variable at the second phase sampling.
$\alpha_i^{+r} = \left[\bar{y}_2 * \prod_{i=1}^{r}\left(\frac{\bar{x}_{(1)i}}{\bar{x}_{(2)i}}\right)^{\alpha_i}\right]$: This is a full information case estimator with first phase sample data available. $i = 1, 2, \cdots, r$. The $+$ symbol before r means that the estimator uses ratio estimation method. However, the presence of $-$ symbol means it is a product estimation method.

$\beta_{2.j}^{-k} = \left[\bar{y}_2 * \prod_{j=r+1}^{k}\left(\frac{\bar{x}_{(2)j}}{\bar{X}_j}\right)^{\beta_j}\right]$: This is a full information case estimator with second phase sample data available. $j = (r+1), (r+2), \cdots, k$. The counter j initializes its counting from where the last counter (i) stopped its counting. The $-$ symbol before superscript k means the estimator uses product estimation method.

$\gamma_f^{+g} = \left[\bar{y}_2 * \prod_{f=1}^{g}\left(\frac{\bar{x}_{(1)f}}{\bar{x}_{(2)f}}\right)^{\gamma_f}\right]$: This is a no information case ratio estimator with $f = 1, 2, \cdots, g$.

3.2 *Proposed mixture ratio estimator in two-phase sampling for Partial Information Case II* (PIC-II)

If our interest is to estimate the population mean for a mixture ratio estimator using multi-auxiliary variables and attributes in two-phase sampling when the population information on the k auxiliary variables are known, population information on the $(k+1)$ to h auxiliary attributes are not known, but the population information on $(h+1)$ to q auxiliary attributes are not known. Then, the mixture ratio estimator is presented as:

$$t_4 = \bar{y}_2\left[\prod_{i=1}^{k}\left(\frac{\bar{x}_{(1)_i}}{\bar{x}_{(2)_i}}\right)^{\alpha_i}\left(\frac{\bar{X}_i}{\bar{x}_{(1)_i}}\right)^{\beta_i}\right]\left[\prod_{f=k+1}^{h}\left(\frac{p_{(1)_f}}{p_{(2)_f}}\right)^{\gamma_f}\left(\frac{P_f}{p_{(1)_f}}\right)^{\lambda_f}\right]\left[\prod_{g=h+1}^{q}\left(\frac{p_{(1)_g}}{p_{(2)_g}}\right)^{\gamma_g}\right] \tag{3.4}$$

The schema for estimator t_4 is presented as:

$$t_4^* = \left(\overbrace{\bar{y}_2 * \underbrace{\alpha_i^{+k} * \beta_{1.i}^{+k}}_{AV(FIC)} * \underbrace{\gamma_f^{+h} * \lambda_{1.f}^{+h} * \gamma_g^{+q}}_{AA(PIC)}}^{Ratio(PIC-II)}\right) \tag{3.5}$$

Applying the equations 2 and 5 to equation 17 yields

$$MSE(t_4) = E_1E_{2/1}\left[\bar{e}_{y_2} + \bar{Y}\sum_{i=1}^{k}\alpha_i\frac{(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i})}{\bar{X}_i} - \bar{Y}\sum_{i=1}^{k}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i}\right.$$
$$\left. + \bar{Y}\sum_{f=k+1}^{h}\gamma_f\frac{(\bar{e}_{\tau(1)_f} - \bar{e}_{\tau(2)_f})}{P_f} - \bar{Y}\sum_{f=k+1}^{h}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f} + \bar{Y}\sum_{g=h+1}^{q}\gamma_g\frac{(\bar{e}_{\tau(1)_g} - \bar{e}_{\tau(2)_g})}{P_g}\right]^2 \tag{3.6}$$

To obtain the optimum values for α_i, β_i, γ_f , λ_f and γ_g , we obtain the partial derivative with respect to α_i, β_i, γ_f , λ_f and γ_g and equating it to zero, hence, solve for the parameters.

$$\alpha_i = \frac{C_y(-1)^{i+1}\left|R_{yx_i}\right|_{y\underline{x}_k}}{C_{xi}\left|R\right|_{\underline{x}_k}} \text{ for } i = 1, 2, \cdots, k \tag{3.7}$$

$$\gamma_f = \frac{C_y(-1)^{f+1}\left|R_{y\tau_f}\right|_{y\underline{\tau}_h}}{C_{\tau_f}\left|R\right|_{\underline{\tau}_h}} \text{ for } f = k+1, k+2, \cdots, h \tag{3.8}$$

$$\gamma_g = \frac{C_y(-1)^{g+1}\Big|R_{y\tau_g}\Big|_{y\underline{\tau}_g}}{C_{\tau_g}\Big|R\Big|_{\underline{\tau}_g}} \text{ for } g = f+1, f+2, \cdots, q \tag{3.9}$$

$$\beta_i = \frac{C_y(-1)^{i+1}\Big|R_{yx_i}\Big|_{y\underline{x}_k}}{C_{xi}\Big|R\Big|_{\underline{x}_k}} \text{ for } i = 1, 2, \cdots, k \tag{3.10}$$

$$\lambda_f = \frac{C_y(-1)^{f+1}\Big|R_{y\tau_f}\Big|_{y\underline{\tau}_h}}{C_{\tau_f}\Big|R\Big|_{\underline{\tau}_h}} \text{ for } f = k+1, k+2, \cdots, h \tag{3.11}$$

Simplifying equation 19 gives

$$\begin{aligned} MSE(t_4) = E_1E_{2/1}\Bigg[\bar{e}_{y_2}\Bigg(\bar{e}_{y_2} + \bar{Y}\sum_{i=1}^{k}\alpha_i\frac{(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i})}{\bar{X}_i} - \bar{Y}\sum_{i=1}^{k}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i} \\ + \bar{Y}\sum_{f=k+1}^{h}\gamma_f\frac{(\bar{e}_{\tau(1)_f} - \bar{e}_{\tau(2)_f})}{P_f} - \bar{Y}\sum_{f=k+1}^{h}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f} + \bar{Y}\sum_{g=h+1}^{q}\gamma_g\frac{(\bar{e}_{\tau(1)_g} - \bar{e}_{\tau(2)_g})}{P_g}\Bigg)\Bigg] \end{aligned} \tag{3.12}$$

Applying expectation to equation 25

$$\begin{aligned} MSE(t_4) = \bar{Y}^2C_y\Bigg[\theta_2C_y + (\theta_1 - \theta_2)\sum_{i=1}^{k}\alpha_iC_{x_i}\rho_{yx_i} - \theta_1\sum_{i=1}^{k}\beta_iC_{x_i}\rho_{yx_i} \\ + (\theta_1 - \theta_2)\sum_{f=k+1}^{h}\gamma_fC_{\tau_f}\rho_{y\tau_f} - \theta_1\sum_{f=k+1}^{h}\lambda_fC_{\tau_f}\rho_{y\tau_f} + (\theta_1 - \theta_2)\sum_{g=h+1}^{q}\gamma_gC_{\tau_g}\rho_{\tau_g}\Bigg] \end{aligned} \tag{3.13}$$

Substitute the optimum equations obtained for α_i, β_i, γ_f, λ_f and γ_g hence, simplify:

$$MSE(t_4) = \bar{Y}^2C_y^2\Big[\theta_2 - \theta_2\rho^2_{y.\underline{x}_k} - \theta_2\rho^2_{y.\underline{\tau}_h} - \theta_2\rho^2_{y.\underline{\tau}_q} + \theta_1\rho^2_{y.\underline{\tau}_q}\Big] \tag{3.14}$$

$$MSE(t_4) = \bar{Y}^2C_y^2\Big[\theta_2\Big(1 - \rho^2_{y.(\underline{x},\underline{\tau})_q}\Big) + \theta_1\rho^2_{y.\underline{\tau}_q}\Big] \tag{3.15}$$

3.3 *Proposed mixture ratio estimator in two-phase sampling for Partial Information Case III ($PIC - III$)*

If our interest is to estimate the population mean for a mixture ratio estimator using multi-auxiliary variables and attributes in two-phase sampling when the population information on the auxiliary variables from $(1 - r)$ are known but for $(r + 1)$ to k are unknown and the population information of the auxiliary attributes from $(k+1)$ to q are known, then we suggest the estimator:

$$t_5 = \bar{y}_2 \Bigg[\prod_{i=1}^{r} \left(\frac{\bar{x}_{(1)_i}}{\bar{x}_{(2)_i}} \right)^{\alpha_i} \left(\frac{\bar{X}_i}{\bar{x}_{(1)_i}} \right)^{\beta_i} \Bigg] \Bigg[\prod_{j=r+1}^{k} \left(\frac{\bar{x}_{(1)_j}}{\bar{x}_{(2)_j}} \right)^{\alpha_j} \Bigg] \Bigg[\prod_{f=K+1}^{q} \left(\frac{p_{(1)_f}}{p_{(2)_f}} \right)^{\gamma_f} \left(\frac{P_f}{p_{(1)_f}} \right)^{\gamma_f} \Bigg] \tag{3.16}$$

The schema for estimator t_5 is presented as:

$$t_5^* = \Bigg[\overbrace{\bar{y}_2 * \underbrace{\alpha_i^{+r} * \beta_{1.i}^{+r} * \alpha_j^{+k}}_{AV(PIC)} * \underbrace{\gamma_f^{+q} * \lambda_{1.f}^{+q}}_{AA(FIC)}}^{Ratio(PIC-II)} \Bigg] \tag{3.17}$$

Applying equation 2 and 5 to equation 29 gives

$$\begin{aligned} MSE(t_5) = E_1 E_{2/1} \Bigg[& \bar{e}_{y_2} + \bar{Y} \sum_{i=1}^{r} \alpha_i \frac{(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i})}{\bar{X}_i} - \bar{Y} \sum_{i=1}^{r} \beta_i \frac{\bar{e}_{x(1)_i}}{\bar{X}_i} \\ & + \bar{Y} \sum_{j=r+1}^{k} \alpha_j \frac{(\bar{e}_{x(1)_j} - \bar{e}_{x(2)_j})}{\bar{X}_j} + \bar{Y} \sum_{f=k+1}^{q} \gamma_f \frac{(\bar{e}_{\tau(1)_f} - \bar{e}_{\tau(2)_f})}{P_f} - \bar{Y} \sum_{f=k+1}^{q} \lambda_f \frac{\bar{e}_{\tau(1)_f}}{P_f} \Bigg]^2 \end{aligned} \tag{3.18}$$

To obtain the optimum values for α_i, β_i, α_j, γ_f and λ_f, we perform the partial derivative with respect to α_i, β_i, α_j, γ_f and λ_f and equate it to zero, hence, solve for the parameters.

$$\alpha_i = \frac{C_y (-1)^{i+1} \Big| R_{yx_i} \Big|_{y\underline{x}_r}}{C_{x_i} \Big| R \Big|_{\underline{x}_r}} \text{ for } i = 1, 2, \cdots, r \tag{3.19}$$

$$\alpha_j = \frac{C_y (-1)^{j+1} \Big| R_{yx_j} \Big|_{y\underline{x}_k}}{C_{x_j} \Big| R \Big|_{\underline{x}_k}} \text{ for } j = r+1, r+2, \cdots, k \tag{3.20}$$

$$\gamma_f = \frac{C_y (-1)^{f+1} \Big| R_{y\tau_f} \Big|_{y\underline{\tau}_q}}{C_{\tau_f} \Big| R \Big|_{\underline{\tau}_q}} \text{ for } f = k+1, k+2, \cdots, q \tag{3.21}$$

$$\beta_i = \frac{C_y (-1)^{i+1} \Big| R_{yx_i} \Big|_{y\underline{x}_r}}{C_{x_i} \Big| R \Big|_{\underline{x}_r}} \text{ for } i = 1, 2, \cdots, r \tag{3.22}$$

$$\lambda_f = \frac{C_y(-1)^{f+1}\left|R_{y\tau_f}\right|_{y\underline{\tau}_q}}{C_{\tau_f}\left|R\right|_{\underline{\tau}_q}} \text{ for } f = k+1, k+2, \cdots, h \tag{3.23}$$

Simplify equation 31 gives

$$MSE(t_5) = E_1E_{2/1}\Bigg[\bar{e}_{y_2}\Bigg(\bar{e}_{y_2} + \bar{Y}\sum_{i=1}^{r}\alpha_i\frac{(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i})}{\bar{X}_i} - \bar{Y}\sum_{i=1}^{r}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i} + \bar{Y}\sum_{j=r+1}^{k}\alpha_j\frac{(\bar{e}_{x(1)_j} - \bar{e}_{x(2)_j})}{\bar{X}_j} + \bar{Y}\sum_{f=k+1}^{q}\gamma_f\frac{(\bar{e}_{\tau(1)_f} - \bar{e}_{\tau(2)_f})}{P_f} - \bar{Y}\sum_{f=k+1}^{q}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f}\Bigg)\Bigg] \tag{3.24}$$

Applying expectation to equation 37 and substitute the optimum equations obtained for α_i, β_i, α_j, γ_f and λ_f, hence, simplify:

$$MSE(t_5) = \bar{Y}^2C_y^2\left[\theta_2 - \theta_2\rho^2_{y.\underline{x}_r} + \theta_1\rho^2_{y.\underline{x}_k} - \theta_2\rho^2_{y.\underline{x}_k} - \theta_2\rho^2_{y.\underline{\tau}_q}\right] \tag{3.25}$$

$$MSE(t_5) = \bar{Y}^2C_y^2\left[\theta_2\left(1 - \rho^2_{y.(\underline{x},\underline{\tau})_q}\right) + \theta_1\rho^2_{y.\underline{x}_k}\right] \tag{3.26}$$

4. Results and discussion

4.1 *Comparison of the estimators in PIC: Case − I, Case − II and Case − III*

4.1.1 Comparison of PIC − I and PIC − II

$$MSE(t_3) > MSE(t_4) \tag{4.1}$$

$$\theta_1\rho^2_{y.\underline{x}_k} - \theta_2\rho^2_{y.\underline{x}_r} > 0 \tag{4.2}$$

Recall that $(\theta_2 > \theta_1)$.

Case I

From equation 11 ($PIC - I$), if the number of auxiliary variables within $i = 1, 2, \cdots, r$ is more than the number of auxiliary variables within $j = (r+1), (r+1), \cdots, k$, then it is expected that $\left(\theta_1\rho^2_{y.\underline{x}_r} > \theta_2\rho^2_{y.\underline{x}_k}\right)$. Hence, equation 41 is false. This implies that estimator in $PIC - I$ is efficient over estimator in $PIC - II$.

Case II

From equation 11 ($PIC - I$), if the number of auxiliary variables within $i = 1, 2, \cdots, r$ is less than the number of auxiliary variables within $j = (r+1), (r+1), \cdots, k$, then it is expected that $\left(\theta_1\rho^2_{y.\underline{x}_r} < \theta_2\rho^2_{y.\underline{x}_k}\right)$. Hence, equation 41 is true. This implies that estimator in $PIC - II$ is efficient over estimator in $PIC - I$.

4.1.2 Comparison of $PIC - I$ and $PIC - III$

$$MSE(t_3) > MSE(t_5) \tag{4.3}$$

$$\theta_1 \rho^2_{y.\underline{\tau}_q} - \theta_2 \rho^2_{y.\underline{\tau}_h} > 0 \tag{4.4}$$

Recall that $(\theta_2 > \theta_1)$

Case I

From equation 11 $(PIC-I)$, if the number of auxiliary attributes within $f = (k+1), (k+1), \cdots, h$ is more than the number of auxiliary attributes within $g = (h+1), (h+1), \cdots, q$, then it is expected that $\left(\theta_1 \rho^2_{y.\underline{\tau}_h} > \theta_2 \rho^2_{y.\underline{\tau}_q}\right)$. Hence, equation 43 is false. This implies that estimator in $PIC - I$ is efficient over estimator in $PIC - III$.

Case II

From equation 11 $(PIC - I)$, if the number of auxiliary attributes within $f = (k+1), (k+1), \cdots, h$ is less than the number of auxiliary attributes within $g = (h+1), (h+1), \cdots, q$, then it is expected that $\left(\theta_1 \rho^2_{y.\underline{\tau}_h} < \theta_2 \rho^2_{y.\underline{\tau}_q}\right)$. Hence, equation 43 is false. This implies that estimator in $PIC - III$ is efficient over estimator in $PIC - I$.

4.1.3 Comparison of PIC-II and PIC-III

$$MSE(t_4) > MSE(t_5) \tag{4.5}$$

$$\left(\theta_1 \rho^2_{y.\underline{\tau}_q} - \theta_2 \rho^2_{y.\underline{\tau}_h}\right) > \left(\theta_1 \rho^2_{y.\underline{x}_k} - \theta_2 \rho^2_{y.\underline{x}_r}\right) \tag{4.6}$$

Recall that $(\theta_2 > \theta_1)$

Case I

If $\left(\rho^2_{y.\underline{\tau}_q} = \rho^2_{y.\underline{x}_k}\right)$ and $\left(\rho^2_{y.\underline{\tau}_h} = \rho^2_{y.\underline{x}_r}\right)$, this implies that $MSE(t_4) = MSE(t_5)$. Hence, $PIC - II$ estimator is equally efficient as $PIC - III$ estimator. However, we subject the use of $PIC - II$ and $PIC - III$ to the available of auxiliary variables and auxiliary attributes.

4.1.4 Comparison of $PIC - II$ and NIC

$$MSE(t_4) < MSE(t_2) \tag{4.7}$$

$$-\theta_1 \rho^2_{y.\underline{x}_k} - \theta_2 \rho^2_{y.\underline{\tau}_h} < 0 \tag{4.8}$$

Recall that $(\theta_1 > 0)$ and $(\theta_2 > 0)$. Similarly, it is expected that $\left(\rho^2_{y.\underline{x}_k} > 0\right)$ and $\left(\rho^2_{y.\underline{\tau}_h} > 0\right)$. Hence, equation 47 is true. This implies that $MSE(t_4) < MSE(t_2)$. Therefore, $PIC - II$ estimator is efficient over NIC estimator.

4.1.5 Comparison of PIC-III *and* NIC

$$MSE(t_5) < MSE(t_2) \tag{4.9}$$

$$-\theta_1\rho^2_{y.\underline{\tau}_q} - \theta_2\rho^2_{y.\underline{x}_r} < 0 \tag{4.10}$$

Recall that ($\theta_1 > 0$) and ($\theta_2 > 0$). Similarly, it is expected that ($\rho^2_{y.\underline{\tau}_q} > 0$) and ($\rho^2_{y.\underline{x}_r} > 0$). Hence, equation 49 is true. This implies that $MSE(t_5) < MSE(t_2)$. Therefore, $PIC - III$ estimator is efficient over NIC estimator.

5. Conclusion

We have proposed two partial information cases ($PIC - I$ and $PIC - II$) in addition to the three estimators established by Waweru *et al.* (2014). The efficiency of our estimators is subjected to some conditions. The efficiency of $PIC - II$ over $PIC - I$ depends on the availability of higher number of auxiliary variables in the FIC over NIC in $PIC-I$ estimator. Similarly, the efficiency of $PIC-III$ estimator over $PIC-I$ estimator depends on the availability of higher number of auxiliary attributes in the FIC over NIC in $PIC - I$ estimator. The condition under which $PIC - II$ is equally efficient as $PIC-III$ was established. However, we subject the use of $PIC-II$ or $PIC-III$ to the availability of auxiliary variables and auxiliary attributes. Finally, our proposed estimators gained efficiency over the NIC estimator that was established by Waweru *et al.* (2014).

References

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

[2] S. Arora and L. Bansi: *New Mathematical Statistics* Satya Prahashar, New Dehli (1989).

[3] S. Bahl and R.K. Tetuja: *Ratio and Product Estimator* Information and Optimization Science. Vol. 12, Pp. 159-163. http://dx.doi.org/10.1080/025522667.1991-10699058 (1991).

[4] W.G. Cochran: *The estimation of the yields of the cereal Experiments by Sampling for the Ratio of Grain to total Produce* Journal of Agricultural Science, 30, 262-275. http://dx.doi.org/10.1017/s0021859600048012 (1940).

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

[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. Vol. 22, Pp. 43-50 (2006).

[7] J. Kungu, G. Chumba and L. Odongo: *Mixture Regression estimators using multi-auxiliary variables and attributes in two-phase sampling* Open Journal of Statistics. Vol. 4, Pp. 355-366. http://dx.doi.org/10.4236/ols.2014.45035 (2014).

[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. Vol. 18, Pp. 1518-1526 (2012).

[9] J. Neyman: *On the two aspects of representing method of stratified sampling and the method of purposive selection* J. Roy. Statist. Soc. 97, 558-606 (1934).

[10] J. Neyman: *Contribution of the theory of sampling human populations* J. Amer. Statist. Assoc. 33, 101-116 (1938).

[11] I. Olkin: *Multivariate ratio estimation for finite population* Biometrika Vl. 45, pp. 154-165. http://dx.doi.org/10.1093/biomet/45.1-2.154 (1958).

[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 M. Hanif: *Estimation of population mean in single phase sampling with or without additional information*, Patistan Journal of Statistics Vol. 23, Pp. 99-118 (2007).

[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, 475 (1937).

[16] P.M. Waweru, J. Kungu and J. Kahiri: *Mixture Ratio Estimators Using Multi-Auxiliary Variables and Attributes for Two-Phase Sampling* Open Journal of Statistics. 4, 776-788. http://dx.doi.org/10.4236/ojs.2014.49073 (2014).