GENERAL CLASS OF RATIO-CUM-PRODUCT ESTIMATORS IN TWO-PHASE SAMPLING USING MULTI-AUXILIARY VARIABLES

Peter I. Ogunyinka [1], Emmanuel Femi Ologunleko [2], Ademola A. Sodipo [2]

[1] Olabisi Onabanjo University, Ago-Iwoye, Nigeria, Department of Mathematical Sciences
[2] University of Ibadan, Ibadan, Nigeria, Department of Statistics

Corresponding Author: Peter I. Ogunyinka, ogunyinka.peter@gmail.com

ABSTRACT: The generalised mixed ratio-cum-product estimators in Two-phase sampling has been developed and recommended using multi-auxiliary variables for Full Information Case (FIC), Partial Information Case I (PIC-I) and No Information Case (NIC) estimators. However, the PIC-I estimators would not obey the conditions of usage if both the ratio and product estimator components were not in partial information case, simultaneously. Hence, this study has considered two additional PIC (PIC-II and PIC-III) estimators which satisfied two more conditions for any PIC estimator in the family mixed estimator. Theoretical comparison established that the proposed PIC-II and PIC-III estimators were asymptotically efficient than PIC-I estimator. Similarly, the empirical analysis and comparison for thirty three simulated populations, following normal distribution, confirmed the asymptotic efficiency of the proposed estimators. The proposed estimators were recommended not as substitute but as complimentary estimators to the PIC-I estimator, subject to the confirmation of the conditions of usage. Special case estimators were developed based on the settings of the unknown constants in the proposed estimators.
KEYWORDS: Ratio-cum-product estimator, Multi-auxiliary variables, Two-phase sampling, Partial information case.

1. INTRODUCTION

The incorporation of auxiliary variables in the estimation of the population parameter has, over the years, proved to be efficient and reliable. ([Gra62]) seems to be the first author to implement auxiliary information in the estimation of English population (human population estimation). High correlation coefficient between the study variable and the auxiliary variable is a significant factor that is necessary in maximizing the advantages of auxiliary variables. Different estimation methods have been developed by survey statisticians that use auxiliary variables for efficient estimator. ([Coc40]) developed ratio estimator in single phase sampling. The presence of positive and high correlation coefficient between the study and auxiliary variables is a primary requirement for the use of ratio estimator. ([Rob57]) proposed the product estimator in single phase sampling. Product estimator requires the presence of negative and high correlation coefficient between the study variable and the auxiliary variables. Regression and difference estimators are other estimators that have been developed alongside with ratio and product estimators. The use of more than one auxiliary variable in the estimation of the study variable has been implemented by ([Raj65]). This method is called multi-auxiliary variable in estimation. ([Moh67]) initiated the use of two estimators (ratio and product) estimators in single phase sampling. This method is tagged mixed estimation method. Ratio-cum-product estimator is an instance of mixed estimator and has been confirmed to improve the efficiency of any estimator. ([Sin67a]) and ([Sin67b]) developed ratio-cum-product estimator which proved efficient than either of ratio and product estimator. Other literatures that have discussed about ratio-cum-product estimator include ([SS78]) and ([TS09]). These estimators were improvement on the single-phase sampling estimators with one on two auxiliary variables.

The use of ratio, regression, product and difference estimators in single phase sampling requires that the population mean of the auxiliary variables should be known prior to the estimation of the study variable. However, when the population total of the auxiliary variable is not known in advance, ([Ney38]) has introduced two-phase sampling or double sampling to be used instead of the single phase sampling. Among survey statisticians, the term two phase sampling has been preferred over double sampling because double sampling is a common and known term among the quality control statisticians ([Kee05]). Two-phase sampling is a sampling scheme and estimation method when it is combined with any of ratio, regression, product or difference estimator. Two phase sampling also uses mixed estimator. ([CS11]) has developed ratio-cum-product estimator in two phase sampling while ([CS14]) has developed mixed ratio-cum-product estimators considering the two cases where second-phase sampling is a dependent sample of the first-phase

(nested two-phase sampling) and the second phase sample is an independent sample of the first phase (non-nested two-phase sampling).
([KN16]) developed improved estimators using ratio-cum-product estimation method with multi-auxiliary variables in two phase sampling. These estimators satisfied the full, no and partial information case methods of utilizing auxiliary variables ([SH07]). These estimators were confirmed to be efficient than either of ratio or product estimators in both single and two phase sampling with one auxiliary variable. However, the partial information case of this family estimator is subjected to the conclusion that there exists partial information about the multi-auxiliary variables, simultaneously, in both the ratio and product estimation components. This partial information case estimator has been identified to be incompatible if the distributions do not, simultaneously, hold partial information case for both ratio and product estimation components in the mixed estimator. This challenge has been identified by ([OS17]) in other estimators. Hence, this study has improved on the partial information case by proposing two additional partial information cases (PIC-II and PIC-III) in addendum to the existing partial information case (PIC-I) in generalised ratio-cum-product estimator in two-phase sampling as proposed by ([KN16]). Similarly, the Mean Square Errors (MSEs) of the proposed estimators for partial information cases have been established theoretical in the subsequent sections.

2. METHODOLOGY

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 (using simple random sampling without replacement), respectively, with $n_1 > n_2$. 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 $\theta_1 < \theta_2$, where θ_1 and θ_2 are variables which values depend on the population and phase-wise sample sizes.
Let $x_{(1)i}$ and $x_{(2)i}$ be the ith auxiliary variable at the first and second phase sampling, respectively and y_2 be the study variable at the second phase sampling.

$$\bar{y}_2 = \left(\bar{Y} + \bar{e}_{y_2}\right);$$
$$\bar{x}_{(1)i} = \left(\bar{X}_i + \bar{e}_{x(1)_i}\right); \qquad (1)$$
$$\bar{x}_{(2)i} = \left(\bar{X}_i + \bar{e}_{x(2)_i}\right); \quad for\ i = 1,2,\dots,p,$$

where $\bar{e}_{y_2}$, $\bar{e}_{x(1)_i}$, $\bar{e}_{x(2)_i}$ are the mean sampling errors and are very small, such that

$$E\left(\bar{e}_{y_2}\right) = E\left(\bar{e}_{x(1)_i}\right) = E\left(\bar{e}_{x(2)_i}\right) = 0.$$

Furthermore, the following phase-wise error conditional expectation results will be useful in obtaining the mean square errors of the proposed estimators later in the work.

$$E_1E_{2/1}\left(\bar{e}_{y_2}\right)^2 = \theta_2\bar{Y}^2C_y^2\ ;$$
$$E_1E_{2/1}\left(\bar{e}_{x(2)_i}\right)^2 = \theta_2\bar{X}_i^2C_{x_i}^2;$$
$$E_1E_{2/1}\left(\bar{e}_{y_2}\bar{e}_{x(2)_i}\right) = \theta_2\bar{Y}\bar{X}_iC_yC_{x_i}\rho_{yx_i};$$
$$E_1E_{2/1}\left(\bar{e}_{x(1)_i}\bar{e}_{x(1)_j}\right) = \theta_1\bar{X}_i\bar{X}_jC_{x_i}C_{x_j}\rho_{x_jx_i}, \qquad for\ i \neq j;$$
$$E_1E_{2/1}\left(\bar{e}_{x(1)_i}\bar{e}_{x(2)_j}\right) = \theta_1\bar{X}_i\bar{X}_jC_{x_i}C_{x_j}\rho_{x_jx_i}, \qquad for\ i \neq j;$$
$$E_1E_{2/1}\left(\bar{e}_{y_2}\bar{e}_{x(1)_i}\right) = \theta_1\bar{Y}\bar{X}_iC_yC_{x_i}\rho_{yx_i};$$
$$E_1E_{2/1}\left(\bar{e}_{y_2}\left(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i}\right)\right) = (\theta_1 - \theta_2)\bar{Y}\bar{X}_iC_yC_{x_i}\rho_{yx_i};$$
$$E_1E_{2/1}\left(\bar{e}_{x(2)_i}\left(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i}\right)\right) = (\theta_1 - \theta_2)\bar{X}_i^2C_{x_i}^2;$$
$$E_1E_{2/1}\left(\bar{e}_{x(1)_i}\left(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i}\right)\right) = 0;$$
$$E_1E_{2/1}\left(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i}\right)^2 = (\theta_2 - \theta_1)\bar{X}_i^2C_{x_i}^2;$$
$$E_1E_{2/1}\left(\left(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i}\right)\left(\bar{e}_{x(1)_j} - \bar{e}_{x(2)_j}\right)\right) = (\theta_2 - \theta_1)\bar{X}_i\bar{X}_jC_{x_i}C_{x_j}\rho_{x_ix_j}, \qquad for\ i \neq j.$$

Finally, according to ([AB89]), the following simplified correlation coefficients matrix will be useful in obtaining the mean square errors for the proposed estimators.

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

2.2 Reviewed Generalized Ratio-cum-Product Estimators in Two-phase Sampling

([KN16]) developed the estimated population mean of a generalized ratio-cum-product estimator in two phase sampling when the population information on all the auxiliary variables are available for both the ratio and product estimation components with the pth multi-auxiliary variables. *This means that the ratio and product estimation components are in full information case (the condition of usage).* The estimator is termed Full Information Case (FIC) estimator and presented as:

$$Q_1 = \bar{y}_2 * \prod_{i=1}^{k}\left(\frac{\bar{X}_i}{\bar{x}_{(2)_i}}\right)^{\alpha_i}\prod_{j=k+1}^{p}\left(\frac{\bar{x}_{(2)_j}}{\bar{X}_j}\right)^{\beta_j}.$$

The corresponding minimized Mean Square Error (MSE) is presented as

$$MSE(Q_1)_{min} \cong \theta_2 \bar{Y}^2 C_y^2\left(1-\rho^2_{y\underset{\sim}{x}_p}\right).$$

Similarly, ([KN16]) developed the estimated population mean of a generalized ratio-cum-product estimator in two phase sampling when the population information on all the auxiliary variables are not available for both the ratio and product estimation components with the pth multi-auxiliary variables. *This means that the ratio and product estimation components are in no information case (the condition of usage).* The estimator is termed No Information Case (NIC) estimator and presented as:

$$Q_2 = \bar{y}_2 * \prod_{i=1}^{k}\left(\frac{\bar{x}_{(1)_i}}{\bar{x}_{(2)_i}}\right)^{\alpha_i}\prod_{j=k+1}^{p}\left(\frac{\bar{x}_{(2)_j}}{\bar{x}_{(1)_i}}\right)^{\beta_j}.$$

The corresponding minimized Mean Square Error (MSE) was given as

$$MSE(Q_2)_{min} \cong \bar{Y}^2 C_y^2\left[\theta_2\left(1-\rho^2_{y\underset{\sim}{x}_p}\right)+\theta_1\rho^2_{y\underset{\sim}{x}_p}\right].$$

Finally, a Partial Information Case (PIC) of the generalized ratio-cum-product estimator was developed by ([KN16]). This estimator concluded that there exists $1,2,\dots,r$ full information case of auxiliary variables and $(r+1),(r+2),\dots,k$ no information case auxiliary variables, all in ratio estimation component. Similarly, it assigns $(k+1),(k+2),\dots,h$ full information case of auxiliary variables and $(h+1),(h+2),\dots,p$ no information case of the auxiliary variables in the product estimation component. *The condition of usage spells that both the ratio and product estimation components are, simultaneously, in partial information case.* This estimator was, also, confirmed to be asymptotically unbiased and consistent estimator. The PIC estimator is presented as:

$$Q_3 = \bar{y}_2\left[\prod_{i=1}^{r}\left(\frac{\bar{x}_{(1)_i}}{\bar{x}_{(2)_i}}\right)^{\alpha_i}\left(\frac{\bar{X}_i}{\bar{x}_{(1)_i}}\right)^{\beta_i}\right]\left[\prod_{j=r+1}^{k}\left(\frac{\bar{x}_{(1)_j}}{\bar{x}_{(2)_j}}\right)^{\alpha_j}\right] * \left[\prod_{m=k+1}^{h}\left(\frac{\bar{x}_{(2)m}}{\bar{x}_{(1)m}}\right)^{\gamma_m}\left(\frac{\bar{x}_{(1)m}}{\bar{X}_i}\right)^{\lambda_m}\right]\left[\prod_{n=h+1}^{p}\left(\frac{\bar{x}_{(2)n}}{\bar{x}_{(1)n}}\right)^{\sigma_n}\right].$$

The corresponding minimized Mean Square Error (MSE) is further simplified as thus:

$$MSE(Q_3)_{min} \cong \bar{Y}^2 C_y^2\left[\theta_2\left(1-\rho^2_{y.\underset{\sim}{x}_p}\right)+\theta_1\left(\rho^2_{y.\underset{\sim}{x}_p}-\rho^2_{y.\underset{\sim}{x}_{r,h}}\right)\right].$$

This work has termed this estimator as Partial Information Case I (PIC-I). This is necessary in order to relate this existing estimator to the proposed estimators. However, these reviewed estimators have been confirmed to be asymptotically efficient. *This work has identified that in application, the condition of using the PIC-I estimator will not be the only satisfactory condition for any PIC. Hence, two new conditions have been identified to proposed two new PIC estimators.* These proposed PIC estimators will be presented in the next section.

2.3 Proposed Generalized Ratio-cum-Product Estimator in Two-Phase Sampling for Partial Information Case II (PIC-II)

This study proposes a partial information case for a generalized ratio-cum-product estimator using multi-auxiliary variables in two phase sampling. It assumes that the ratio estimation component has $1,2,\dots,k$ auxiliary variables all in full information case. The product estimation component has $(k+1),(k+2),\dots,h$ auxiliary variables in full information case but $(h+1),(h+2),\dots,p$ auxiliary variables in no information case. *The condition of usage spells that the ratio estimation component is in FIC while the product estimation component is in PIC (that is, not simultaneously).* This is a partial information case for the product estimation component. The combination of the ratio and product estimation components forms the general class of partial information case tagged as Partial Information Case II (PIC-II). It is presented as

$$Q_4 = \bar{y}_2\left[\prod_{i=1}^{k}\left(\frac{\bar{x}_{(1)_i}}{\bar{x}_{(2)_i}}\right)^{\alpha_i}\left(\frac{\bar{X}_i}{\bar{x}_{(1)_i}}\right)^{\beta_i}\right] * \left[\prod_{m=k+1}^{h}\left(\frac{\bar{x}_{(2)_m}}{\bar{x}_{(1)_m}}\right)^{\gamma_m}\left(\frac{\bar{x}_{(1)_m}}{\bar{X}_m}\right)^{\lambda_m}\right]\left[\prod_{n=h+1}^{p}\left(\frac{\bar{x}_{(2)_n}}{\bar{x}_{(1)_n}}\right)^{\sigma_n}\right] \quad (2)$$

([OS17]) had developed abridged version of a lengthy estimator in two-phase sampling estimators without the estimator losing its value and meaning. The estimator schema for equation (2) is presented as

$$Q_4^* = \left\{ \overbrace{\bar{y}_2 * \underbrace{\alpha_i^{+k} * \beta_{1.i}^{+k}}_{\substack{Ratio\ (FIC) \\ AV}} * \underbrace{\gamma_m^{-h} * \lambda_{1.m}^{-h} * \sigma_n^{-p}}_{\substack{Product\ (PIC) \\ AV}}}^{Ratio-cum-Product\ (PIC-II)} \right\} \quad (3)$$

On substituting equation (1) into equation (2), we have

$$MSE(Q_4) = E_1E_{2/1}\left[\bar{e}_{y2} + \bar{Y}\sum_{i=1}^{k}\alpha_i\frac{(\bar{e}_{x(1)_i}-\bar{e}_{x(2)_i})}{\bar{X}_i} - \bar{Y}\sum_{i=1}^{k}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i} - \bar{Y}\sum_{m=k+1}^{h}\gamma_m\frac{(\bar{e}_{x(1)_m}-\bar{e}_{x(2)_m})}{\bar{X}_m} + \bar{Y}\sum_{m=k+1}^{h}\lambda_m\frac{\bar{e}_{x(1)m}}{\bar{X}_m} - \bar{Y}\sum_{n=h+1}^{p}\sigma_n\frac{(\bar{e}_{x(1)_n}-\bar{e}_{x(2)_n})}{\bar{X}_n}\right]^2.$$

Next is to apply Taylors series expansion and partial differentiation with respect to the five unknown constants ($\alpha_i, \beta_i, \gamma_m, \lambda_m\ and\ \sigma_n$) and make each of the constants a subject formula. Hence, the optimum values of the unknown constants are obtained as

$$\alpha_i = \frac{C_y(-1)^{i+1}|R_{yx_i}|_{y\underset{\sim}{x}_k}}{C_{x_i}|R|_{\underset{\sim}{x}_k}},\ for\ i = 1,2,\dots,k;$$

$$\gamma_m = \frac{-C_y(-1)^{m+1}|R_{yx_m}|_{y\underset{\sim}{x}_h}}{C_{x_m}|R|_{\underset{\sim}{x}_h}},\quad for\ m = k+1, k+2, \dots, h;$$

$$\beta_i = \frac{C_y(-1)^{i+1}|R_{yx_i}|_{y\underset{\sim}{x}_k}}{C_{x_i}|R|_{\underset{\sim}{x}_k}},\ for\ i = 1,2,\dots,k;$$

$$\sigma_n = \frac{-C_y(-1)^{n+1}|R_{yx_n}|_{y\underset{\sim}{x}_p}}{C_{x_n}|R|_{\underset{\sim}{x}_p}},\quad for\ n = h+1, h+2, \dots, p;$$

$$\lambda_m = \frac{-C_y(-1)^{m+1}|R_{yx_m}|_{y\underset{\sim}{x}_h}}{C_{x_m}|R|_{\underset{\sim}{x}_h}}\quad for\ m = k+1, k+2, \dots, h.$$

Finally, we substitute the obtained constants and apply the conditional expectation to obtain the minimized MSE for the proposed estimator Q_4

$$MSE(Q_4)_{min} \cong \bar{Y}^2 C_y^2\left[\theta_2\left(1-\rho_{y.\underset{\sim}{x}_p}^2\right) + \theta_1\left(\rho_{y.\underset{\sim}{x}_p}^2 - \rho_{y.\underset{\sim}{x}_{k,h}}^2\right)\right] \quad (4)$$

2.4 Proposed Generalized Ratio-cum-Product Estimator in Two-Phase Sampling for Partial Information Case III (PIC-III)

This estimator proposes a partial information case for a generalized ratio-cum-product estimator using multi-auxiliary variables in two phase sampling. It assumes that there is $1,2,\dots,r$ auxiliary variables with full information about the auxiliary variables while $(r+1),(r+2),\dots,k$ with no information about the auxiliary variables all in ratio estimation component. This is a partial information case for the ratio estimation component. Finally, it assumes $(k+1),(k+2),\dots,p$ with full information only in the product estimation component. *The condition of usage spells that the ratio estimation component is in PIC while the product estimation component is in FIC (that is, not simultaneously).* The estimated population mean of the study variable combining the ratio and product estimation components is tagged Partial Information Case III (PIC-III) and presented as

$$Q_5 = \bar{y}_2 * \left[\prod_{i=1}^{r}\left(\frac{\bar{x}_{(1)_i}}{\bar{x}_{(2)_i}}\right)^{\alpha_i}\left(\frac{\bar{X}_i}{\bar{x}_{(1)_i}}\right)^{\beta_i}\right]\left[\prod_{j=r+1}^{k}\left(\frac{\bar{x}_{(1)_j}}{\bar{x}_{(2)_j}}\right)^{\alpha_j}\right] * \left[\prod_{m=k+1}^{p}\left(\frac{\bar{x}_{(2)_m}}{\bar{x}_{(1)_m}}\right)^{\gamma_m}\left(\frac{\bar{x}_{(1)_m}}{\bar{X}_m}\right)^{\lambda_m}\right] \quad (5)$$

The schema for estimator Q_5 is presented as:

$$Q_5^* = \left\{ \overbrace{\bar{y}_2 * \underbrace{\alpha_i^{+r} * \beta_{1.i}^{+r} * \alpha_j^{+k}}_{\substack{Ratio\ (PIC) \\ AV}} * \underbrace{\gamma_m^{-p} * \lambda_{1.m}^{-p}}_{\substack{Product\ (FIC) \\ AV}}}^{Ratio-cum-Product\ (PIC-III)} \right\} \quad (6)$$

and on substituting equation (1) into equation (5), we have

$$\boldsymbol{MSE(Q_5) = E_1E_{2/1}\left[\bar{e}_{y2} + \bar{Y}\sum_{i=1}^{r}\alpha_i\frac{(\bar{e}_{x(1)_i}-\bar{e}_{x(2)_i})}{\bar{X}_i} - \bar{Y}\sum_{i=1}^{r}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i} + \bar{Y}\sum_{j=r+1}^{k}\alpha_j\frac{(\bar{e}_{x(1)_j}-\bar{e}_{x(2)_j})}{\bar{X}_j} - \bar{Y}\sum_{m=k+1}^{p}\gamma_m\frac{(\bar{e}_{x(1)_m}-\bar{e}_{x(2)_m})}{\bar{X}_m} + \bar{Y}\sum_{m=k+1}^{p}\lambda_m\frac{\bar{e}_{x(1)m}}{\bar{X}_m}\right]^2}.$$

Again, we apply Taylors series expansion and partial differentiation with respect to the five unknown constants ($\boldsymbol{\alpha_i, \beta_i, \alpha_j, \gamma_m\ and\ \lambda_m}$) and make each of the constants a subject formula. Hence, the optimum values of the unknown constants are obtained as

$$\boldsymbol{\alpha_i = \frac{C_y(-1)^{i+1}|R_{yx_i}|_{y\underset{\sim}{x}_r}}{C_{x_i}|R|_{\underset{\sim}{x}_r}}\qquad for\ i = 1,2,\dots,r;}$$

$$\boldsymbol{\beta_i = \frac{C_y(-1)^{i+1}|R_{yx_i}|_{y\underset{\sim}{x}_r}}{C_{x_i}|R|_{\underset{\sim}{x}_r}}\qquad for\ i = 1,2,\dots,r;}$$

$$\boldsymbol{\alpha_j = \frac{C_y(-1)^{j+1}\left|R_{yx_j}\right|_{y\underset{\sim}{x}_k}}{C_{x_j}|R|_{\underset{\sim}{x}_k}} \; for\; j = r+1, r+2, \ldots, k;}$$

$$\boldsymbol{\gamma_m = \frac{-C_y(-1)^{m+1}\left|R_{yx_m}\right|_{y\underset{\sim}{x}_p}}{C_{x_m}|R|_{\underset{\sim}{x}_p}} \; for\; m = k+1, k+2, \ldots, p;}$$

$$\boldsymbol{\lambda_m = \frac{-C_y(-1)^{m+1}\left|R_{yx_m}\right|_{y\underset{\sim}{x}_p}}{C_{x_m}|R|_{\underset{\sim}{x}_p}} \; for\; m = k+1, k+2, \ldots, p.}$$

Expanding the brackets further and ignoring second and higher order degrees to give:

$$MSE(Q_5) = E_1E_{2/1}\left[\bar{e}_{y2}\left(\bar{e}_{y2} + \bar{Y}\sum_{i=1}^{r}\alpha_i \frac{\left(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i}\right)}{\bar{X}_i} - \bar{Y}\sum_{i=1}^{r}\beta_i \frac{\bar{e}_{x(1)_i}}{\bar{X}_i} + \bar{Y}\sum_{j=r+1}^{k}\alpha_j \frac{\left(\bar{e}_{x(1)_j} - \bar{e}_{x(2)_j}\right)}{\bar{X}_j} - \bar{Y}\sum_{m=k+1}^{p}\gamma_m \frac{\left(\bar{e}_{x(1)_m} - \bar{e}_{x(2)_m}\right)}{\bar{X}_m} + \bar{Y}\sum_{m=k+1}^{p}\lambda_m \frac{\bar{e}_{x(1)m}}{\bar{X}_m}\right)\right].$$

On substituting the obtained constants and applying the conditional expectation, we have obtain the minimized MSE for the proposed estimator Q_5 as

$$MSE(Q_5)_{min} \cong \bar{Y}^2 C_y^2\left[\theta_2\left(1 - \rho^2_{y.\underset{\sim}{x}_p}\right) + \theta_1\left(\rho^2_{y.\underset{\sim}{x}_p} - \rho^2_{y.\underset{\sim}{x}_{r,p}}\right)\right] \quad (7)$$

3. RESULT

3.1 Theoretical Comparison of the Estimators in PIC-I, PIC-II, PIC-III and NIC

Theoretical and empirical comparison of the proposed estimators with the reviewed estimators has been conducted in this section.

3.1.1 Theoretical Comparison of PIC-I and PIC-II:

$$MSE(Q_3)_{min} - MSE(Q_4)_{min} = \theta_1\bar{Y}^2 C_y^2\left(-\rho^2_{y.\underset{\sim}{x}_{r,h}} + \rho^2_{y.\underset{\sim}{x}_{k,h}}\right).$$

It is expected that $(r < k)$ or $\left(\rho^2_{y.\underset{\sim}{x}_{r,h}} < \rho^2_{y.\underset{\sim}{x}_{k,h}}\right)$, hence, $\theta_1\bar{Y}^2 C_y^2\left(-\rho^2_{y.\underset{\sim}{x}_{r,h}} + \rho^2_{y.\underset{\sim}{x}_{k,h}}\right) > 0$. Therefore, Q_4 will be efficient than Q_3.

3.1.2 Theoretical Comparison of PIC-I and PIC-III:

$$MSE(Q_3)_{min} - MSE(Q_5)_{min} = \theta_1\bar{Y}^2 C_y^2\left(-\rho^2_{y.\underset{\sim}{x}_{r,h}} + \rho^2_{y.\underset{\sim}{x}_{r,p}}\right).$$

It is expected that $(h < p)$ or $\left(\rho^2_{y.\underset{\sim}{x}_{r,h}} < \rho^2_{y.\underset{\sim}{x}_{r,p}}\right)$, hence, $\theta_1\bar{Y}^2 C_y^2\left(-\rho^2_{y.\underset{\sim}{x}_{r,h}} + \rho^2_{y.\underset{\sim}{x}_{r,p}}\right) > 0$. Therefore, Q_5 is efficient than Q_3.

3.1.3 Theoretical Comparison of PIC-II and PIC-III:

$$MSE(Q_4)_{min} - MSE(Q_5)_{min} = \theta_1\bar{Y}^2 C_y^2\left(-\rho^2_{y.\underset{\sim}{x}_{k,h}} + \rho^2_{y.\underset{\sim}{x}_{r,p}}\right).$$

Q_4 will be efficient than Q_5 if and only if $[(k > r)\ and\ (h > p)]$ or $\left(\rho^2_{y.\underset{\sim}{x}_{k,h}} < \rho^2_{y.\underset{\sim}{x}_{r,p}}\right)$. However, it is expected that $(k > r)$ but $(h < p)$. Hence, the efficiency between PIC-II and PIC-III could not be determined in the theoretical comparison. The efficiency will be determined in the numerical comparison.

3.1.4 Theoretical Comparison of PIC-II and NIC:

$$MSE(Q_4)_{min} - MSE(Q_2)_{min} = -\theta_1\bar{Y}^2 C_y^2\rho^2_{y.\underset{\sim}{x}_{k,h}} < 0.$$

Since $\left(-\theta_1\bar{Y}^2 C_y^2\rho^2_{y.\underset{\sim}{x}_{k,h}} < 0\right)$, then Q_4 is unconditionally efficient than Q_2.

3.1.5 Theoretical Comparison of PIC-III and NIC:

$$MSE(Q_5)_{min} - MSE(Q_2)_{min} = -\theta_1\bar{Y}^2 C_y^2\rho^2_{y.\underset{\sim}{x}_{r,p}} < 0.$$

Since $\left(-\theta_1\bar{Y}^2 C_y^2\rho^2_{y.\underset{\sim}{x}_{r,p}} < 0\right)$, then Q_5 is unconditionally efficient than Q_2.

3.2 Empirical Comparison of all Estimators

Tables 1 through 8 display the results of the numerical analysis for thirty three simulated populations.

3.3 Special Cases of the proposed Estimators:

This section shows some special cases for the proposed estimators subject to the setting of the unknown constants in each of the two proposed estimators. Tables 7 and 8 present the special case estimators.

4. DISCUSSION AND CONCLUSIONS

This study uses R statistical programming language for the simulation and empirical analysis in the comparison of the five estimators (FIC, NIC, PIC-I, PIC-II and PIC-III). Simulation, following normal distribution, was conducted for thirty three populations at different population sizes, first-phase sample sizes, second-phase sample sizes and seed values for the thirty three simulated populations.

Table 1. Computation of the population sizes, sample sizes, coefficient of variation and correlation coefficient for the numerical analyses

Pop.	N	n_1	n_2	$Seed$	$\bar{Y}^2$	C_y^2	$\rho^2_{y.\underset{\smile}{x}_{6,6}}$	$\rho^2_{y.\underset{\smile}{x}_{3,3}}$	$\rho^2_{y.\underset{\smile}{x}_{6,3}}$	$\rho^2_{y.\underset{\smile}{x}_{3,6}}$
1	10000	3333	1111	385	130089.10	0.0061	0.9794	0.5031	0.6070	0.8797
2	9700	3233	1078	411	129936.17	0.0065	0.9804	0.5149	0.5943	0.8896
3	9400	3133	1044	87	129748.32	0.0060	0.9794	0.4561	0.5900	0.8673
4	9100	3033	1011	72	130396.00	0.0062	0.9797	0.4886	0.5993	0.8857
5	8800	2933	978	227	130222.83	0.0062	0.9805	0.4852	0.5793	0.8883
6	8500	2833	944	224	129726.64	0.0071	0.9834	0.5132	0.6117	0.8926
7	8200	2733	911	79	130331.30	0.0062	0.9812	0.5140	0.5963	0.8782
8	7900	2633	878	30	129431.13	0.0064	0.9807	0.4720	0.5654	0.8898
9	7600	2533	844	159	130204.68	0.0064	0.9803	0.4754	0.5821	0.8764
10	7300	2433	811	121	131093.10	0.0062	0.9812	0.5208	0.6075	0.8759
11	7000	2333	778	354	131231.12	0.0069	0.9828	0.5364	0.6289	0.8933
12	6700	2233	744	267	130842.64	0.0060	0.9808	0.4690	0.5888	0.8756
13	6400	2133	711	74	129541.27	0.0060	0.9793	0.4724	0.5910	0.8694
14	6100	2033	678	307	130653.46	0.0067	0.9816	0.4974	0.5990	0.8925
15	5800	1933	644	29	129252.53	0.0068	0.9818	0.5108	0.6046	0.9049
16	5500	1833	611	43	130357.63	0.0066	0.9819	0.5035	0.6200	0.8786
17	5200	1733	578	0	131805.43	0.0065	0.9820	0.5506	0.6403	0.8987
18	4900	1633	544	176	130017.25	0.0062	0.9804	0.5284	0.6063	0.8870
19	4600	1533	511	480	129121.38	0.0069	0.9806	0.5542	0.6328	0.8998
20	4300	1433	478	47	129150.01	0.0056	0.9795	0.4774	0.5305	0.8858
21	4000	1333	444	192	130379.55	0.0061	0.9815	0.5032	0.5816	0.8730
22	3700	1233	411	295	128875.76	0.0064	0.9828	0.5445	0.6393	0.8819
23	3400	1133	378	418	130405.06	0.0070	0.9829	0.5405	0.6456	0.8902
24	3100	1033	344	176	129369.87	0.0059	0.9785	0.5236	0.6057	0.8786
25	2800	933	311	332	130860.15	0.0058	0.9787	0.4431	0.5439	0.8615
26	2500	833	278	203	130253.48	0.0071	0.9838	0.5520	0.6278	0.9116
27	2200	733	244	150	132305.12	0.0067	0.9837	0.4420	0.6153	0.8930
28	1900	633	211	708	131683.00	0.0058	0.9813	0.4710	0.5968	0.8759
29	1600	533	178	367	129054.51	0.0070	0.9814	0.5239	0.5929	0.8943
30	1300	433	144	328	131713.55	0.0090	0.9894	0.6258	0.6732	0.9280
31	1000	333	111	109	131128.81	0.0079	0.9826	0.5771	0.6435	0.9043
32	700	233	78	11	129516.94	0.0076	0.9876	0.5558	0.6406	0.9193
33	400	133	44	167	129469.12	0.0045	0.9833	0.5395	0.6026	0.8920

Table 2. The Mean Square Errors (MSEs) for the thirty-three simulated populations

Estimator	1	2	3	4	5	6	7	8	9	10	11
FIC	0.0130	0.0132	0.0145	0.0149	0.0142	0.0137	0.0151	0.0159	0.0179	0.0169	0.0172
PIC-I	0.0885	0.0918	0.0963	0.0947	0.0943	0.0997	0.0902	0.1414	0.1434	0.1298	0.1389
PIC-II	0.0720	0.0784	0.0753	0.0767	0.0791	0.0817	0.0770	0.1183	0.1169	0.1085	0.1137
PIC-III	0.0288	0.0285	0.0320	0.0301	0.0291	0.0303	0.0317	0.0383	0.0438	0.0427	0.0416
NIC	0.1681	0.1786	0.1675	0.1742	0.1728	0.1936	0.1728	0.2578	0.2615	0.2575	0.2851

Table 3. The Mean Square Errors (MSEs) for the thirty-three simulated populations (continuation)

Estimator	12	13	14	15	16	17	18	19	20	21	22
FIC	0.0182	0.0210	0.0209	0.0225	0.0233	0.0230	0.0255	0.0296	0.0283	0.0292	0.0311
PIC-I	0.1397	0.1398	0.1480	0.1473	0.1876	0.1698	0.1723	0.1826	0.2106	0.2179	0.2115
PIC-II	0.1112	0.1120	0.1213	0.1225	0.1476	0.1393	0.1470	0.1544	0.1913	0.1870	0.1725
PIC-III	0.0432	0.0468	0.0443	0.0429	0.0588	0.0513	0.0558	0.0586	0.0623	0.0720	0.0727
NIC	0.2510	0.2504	0.2785	0.2827	0.3606	0.3572	0.3440	0.3815	0.3840	0.4164	0.4356

Table 4. The Mean Square Errors (MSEs) for the thirty-three simulated populations (continuation)

Estimator	23	24	25	26	27	28	29	30	31	32	33
FIC	0.0372	0.0428	0.0470	0.0480	0.0521	0.0604	0.0836	0.0782	0.1435	0.1392	0.1963
PIC-I	0.2779	0.2520	0.3322	0.3677	0.4850	0.4920	0.6181	0.7271	0.9796	1.3718	1.4872
PIC-II	0.2207	0.2142	0.2785	0.3116	0.3465	0.3856	0.5375	0.6425	0.8427	1.1297	1.3037
PIC-III	0.0876	0.0888	0.1094	0.1014	0.1246	0.1495	0.1854	0.1878	0.3050	0.3341	0.4618
NIC	0.5719	0.4927	0.5682	0.7764	0.8381	0.8903	1.2303	1.8439	2.1695	2.9585	3.0566

Table 5. Ranking of the Mean Square Errors (MSEs) for the thirty three populations

Info. Case	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
FIC	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*
PIC-I	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*
PIC-II	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*
PIC-III	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*
NIC	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*

Table 6. Ranking of the Mean Square Errors (MSEs) for the thirty three populations (continuation)

Info. Case	22	23	24	25	26	27	28	29	30	31	32	33	Rank Average	Ave. Rank
FIC	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1.00*	*1*
PIC-I	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4*	*4.00*	*4*
PIC-II	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3.00*	*3*
PIC-III	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*2.00*	*2*
NIC	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5.00*	*5*

Table 7. Parameter settings for Partial Information Case II (PIC-II)

SN	α_i	β_i	γ_m	λ_m	σ_n	Partial Information Case II (PIC-II)	Description
01	0	0	1	1	1	$Q_4 = \bar{y}_2 \left[\prod_{m=k+1}^{h} \left(\frac{\bar{x}_{(2)_m}}{\bar{x}_{(1)_m}}\right)^{\gamma_m} \left(\frac{\bar{x}_{(1)_m}}{\bar{X}_m}\right)^{\lambda_m}\right] \left[\prod_{n=h+1}^{p} \left(\frac{\bar{x}_{(2)_n}}{\bar{x}_{(1)_n}}\right)^{\sigma_n}\right]$	General Product Estimators using Multi-Auxiliary Variables in Two-Phase Sampling in PIC-II.
02	1	1	0	0	0	$Q_4 = \bar{y}_2 \left[\prod_{i=1}^{k} \left(\frac{\bar{x}_{(1)_i}}{\bar{x}_{(2)_i}}\right)^{\alpha_i} \left(\frac{\bar{X}_i}{\bar{x}_{(1)_i}}\right)^{\beta_i}\right]$	General Ratio Estimators using Multi-Auxiliary Variables in Two-Phase Sampling in PIC-II.

Table 8. Parameter settings for Partial Information Case III (PIC-III)

SN	α_i	β_i	α_j	γ_m	λ_m	Partial Information Case III (PIC-III)	
01	0	0	1	1	1	$Q_5 = \bar{y}_2 \left[\prod_{j=r+1}^{k} \left(\frac{\bar{x}_{(1)_j}}{\bar{x}_{(2)_j}}\right)^{\alpha_j}\right]\left[\prod_{m=k+1}^{p} \left(\frac{\bar{x}_{(2)_m}}{\bar{x}_{(1)_m}}\right)^{\gamma_m} \left(\frac{\bar{x}_{(1)_m}}{\bar{X}_m}\right)^{\lambda_m}\right]$	General Product Estimators using Multi-Auxiliary Variables in Two-Phase Sampling in PIC-III.
02	1	1	0	0	0	$Q_5 = \bar{y}_2 * \left[\prod_{i=1}^{r} \left(\frac{\bar{x}_{(1)_i}}{\bar{x}_{(2)_i}}\right)^{\alpha_i} \left(\frac{\bar{X}_i}{\bar{x}_{(1)_i}}\right)^{\beta_i}\right]$	General Ratio Estimators using Multi-Auxiliary Variables in Two-Phase Sampling in PIC-III.

Tables 2 through 4 show the computed Mean Square Errors (MSEs) obtained for the thirty three populations and each of the five proposed estimators. Tables 5 and 6 show the ranking of the computed MSEs for the thirty three populations, the average of the ranked values were computed and the ranking of the average values were obtained to decide on the ranking of the proposed estimators. The source code used for this empirical analysis is hosted on github repository as Free and Open Source (FOS) code with the bitly Uniform Resource Locator (URL) https://bit.ly/2JxDb33 .

Tables 2 through 4 reveal that increment in the population and sample sizes leads to decrement in the MSEs for the thirty three populations and for the five estimators. This analysis establishes that the five estimators are asymptotically efficient. It is, also, osbserved that the proposed estimators (PIC-II and PIC-III) have smaller MSEs compared to the reviewed PIC-I estimator. This confirms that PIC-II and PIC-III estimators are asymptotically efficient than PIC-I estimator in the family of the PIC estimators.

Tables 5 and 6 show the ranking of the computed MSEs, the average of the ranked values and the rank of the average values for the thirty three simulated populations and all of the five estimators. The average rank was used for decision. It is observed that the rank of the FIC estimator remains one (1) throughout for the thirty three populations. Similarly, the ranks of the NIC, PIC-I, PIC-II and PIC-III estimators remain fixed throughout the thirty three populations. Finally, the rank of the average ranks establishes that FIC and NIC estimators are best and least ranked estimators, respectively. It, also, shows that the proposed estimators (PIC-II and PIC-III) prove efficient than the reviewed PIC-I estimator. This is, also, confirmed in the theoretical comparison. This argument will remain true based on the correlation coefficient obtained in the empirical analysis. Finally, tables 7 and 8 show some of the special case estimators that were extracted from PIC-II and PIC-III estimators based on the setting of the unknown constants in each estimator.

In conclusion, the proposed PIC-II and PIC-III estimators are asymptotically efficient than the reviewed PIC-I estimator. These proposed estimators have created alternative PIC estimator based on the confirmation of the conditions of usage. The proposed PIC-II and PIC-III estimators do not serve as replacement to the reviewed PIC-I estimator because they have different conditions of usage. Hence, each one of these three PIC estimators should be used based on the stated conditions of usage. Therefore, the proposed estimators are recommended subject to the confirmation of the conditions of usage.

ACKNOWLEDGEMENT

The authors appreciate the funding received from World Bank Trust Fund for Statistical Capacity Building (WB TFSCB) and International Statistical Institute (ISI) to have presented this work in the 62nd International Statistical Institute (ISI) World Statistics Congress held in Kuala Lumpur Convention Centre (KLCC) in Malaysia between August 18 and 23, 2019.

REFERENCES

[AB89] **S. Arora, L. Bansi** – *New Mathematical Statistics*. New Dehli: Satya Prahashar, 1989.

[Coc40] **W. Cochran** – *The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce*. J. Agricultural Sc., 30, 262-275, 1940.

[CS11] **S. Choudhury, B. Singh** – *An improved class of ratio-Cum-Product Estimator of finite population mean in simple random sampling.* International Journal of

Statistics and Analysis, 1(4), 393-403, 2011.

[CS14] **W. W. Chanu, B. K. Singh** – *Improved class of ratio-Cum-Product Estimators of finite population mean in two phase sampling.* Global Journal of Science Frontier Research: Mathematics and Decision Sciences, 14(2), 69-81, 2014.

[Gra62] **J. Graunt** – *Natural and Political Observations upon the Bills of Mortality*. *London*, 1662.

[Kee05] **K. J. Keen** – *Two Phase Sampling*. Encyclopedia of Bio-Statstics, 8, 1-4, 2005.

[KN16] **J. Kung'u, J. Nderitu** –*Generalized Ratio-Cum-Product Estimators for Two-Phase Sampling Using Multi-Auxiliary Variables*. Open Journal of Statistics, 6, 616-627, 2016.

[Moh67] **S. Mohanty** – *Combination of Regression and Ratio Estimate.* Jour. Ind. Statist. Assoc., 5, 16-19, 1967.

[Ney38] **J. Neyman** – *Contribution to the theory of sampling human populations.* J. Amer. Statist. Assoc., 33, 101-116, 1938.

[OS17] **P. I. Ogunyinka, A. A. Sodipo** – *Extension of information Case on Mixture Regression Estimator using Multi-Auxiliary Variables and Attributes on Two-phase Sampling.* Journal of Statistics: Advance in Theory and Applications, 17(1), 1-20, 2017.

[Raj65] **D. Raj** – *On a method of using multi-auxiliary information in Samplele Surveys*. J. Amer. Statist. Assoc., 60, 270-277, 1965.

[Rob57] **D. Robson** – *Application of Multivariate Polykays to the Theory of unbiade Ratio Type Estimators.* Journal of American Statistical Association, 52, 511-522, 1957.

[Sin67a] **M. Singh** – *Ratio cum product method of estimation*. Metrika, 12(1), 34-43, 1967.

[Sin67b] **M. Singh** – *On the estimation of ratio and product of population parameters*. Sankhya, Ser. C, 27, 321-328, 1967.

[SH07] **M. Samiuddin, M. Hanif** – *Estimation of population mean in single phase sampling with or without additional information.* Pakistan Journal of Statistics, 23, 99-118, 2007.

[SS78] **S. Shah, D. Shah** – *Ratio-Cum-Product Estimators for estimating ratio (product) of two population paramaters*. Sankhya, 40, 156-166, 1978.

[TS09] **R. Taylor, B. A. Sharma** – *Modified ratio-cum-product estimator of finite population mean using know Coefficient of variation and coefficient of kurtosis*. Statistics in Transition-New Series. Jul-09, 10(1), 15-24, 2009.