Transactions of the Nigerian Association of Mathematical Physics
Volume 4, (July, 2017), pp237– 250
© Trans. of NAMP

Extension of Information Cases On Mixture Ratio Estimators Using Multi-Auxiliary Variables And Attributes In Two-Phase Sampling

Ogunyinka Peter I.[1] and Sodipo A. A.[2]

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

Abstract

The use of multiple auxiliary characteristics like auxiliary attributes and variables has been confirmed to improve the efficiency of estimators in two-phase sampling. The three ways of utilizing auxiliary characteristics (full, partial and no information cases) have provided flexibility in the usage of these auxiliary characteristics. Literature has developed mixture ratio estimator for Partial Information Case I (PIC-I) considering both auxiliary characteristics as PIC simultaneously in two-phase sampling. However, the estimator becomes unnecessary if there is only PIC in either of the two auxiliary characteristics. This article has proposed two PIC (PIC-II and PIC-III) estimators considering either one of the auxiliary characteristics as PIC at each point in time in two-phase sampling. Theoretical and empirical comparison of the estimators showed that the proposed partial information case estimators (PIC-II and PIC-III) are efficient over estimators in PIC-I and No Information Case (NIC). R statistical software was used in the empirical analysis. Finally, the proposed estimator schema was able to abridge lengthy estimators.

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

1.0. 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 information. Two-phase sampling, among other sampling techniques, maximizes the advantages of auxiliary information (utilization of auxiliary information). The use of auxiliary information at the pre-selection stage was initiated in [1, 2] while Cochran [3] was the first to coin 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 estimation stage. 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 was pioneered in [4].

The works of [5, 6] were summarized in [7] 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 all the parameters of auxiliary variables may be known (called Full Information Case). The third information case called Partial Information Case (PIC) was introduced in [8]. This is the combination of both full information and no information cases into ratio and regression estimation methods.

Correspondence Author: Ogunyinka P.I., Email: ogunyinka.peter@gmail.com, Tel: +2348052003783

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 characteristic about the population could be dichotomous property (present or absent) which is also, highly correlated with the study variable. The use of such dichotomous characteristic, called auxiliary attribute, in Sample Survey has revealed improvement on the estimation of the study variable. 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 include [9, 10, 11, 12, 13]. Mixture estimator uses the combination of auxiliary attributes and variables towards the improvement of an estimator. 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) was proposed in [14]. It was ascertained that estimators developed in [14] gained efficiency over the reviewed estimators of [12].

Mixture regression estimation in two-phase sampling by assuming partial information case in both auxiliary attributes and variables was established in [15]. However, this Partial Information Case I (PIC-I) will not be compatible in the situation where there is no partial information case in both the auxiliary attributes and variables simultaneously. The solution to this challenge was provided in [16] in the proposed partial information cases (PIC-II and PIC-III) by assuming partial information case in auxiliary attributes or variables non-simultaneously.

Similarly, partial information case (PIC-I) for mixture ratio estimator in two-phase sampling was developed in [14]. It is ascertained that this estimator has the same challenge as that of [15]. Hence, this article employs the method used in [16] to establish two additional partial information cases (PIC-II and PIC-III). These proposed estimators will be used when there is partial information case state in either the auxiliary attributes or variables of the concerned estimator.

Finally, the mean square errors of the proposed estimators were established following [17] approach of presenting mean square error.

It is observed that with the mixture use of auxiliary attribute and variable in an estimator, such estimator becomes lengthy to express and may affect the understanding of such estimator. Hence, estimator schema is introduced to abridge any lengthy estimator for full, partial and no information cases.

2.0 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 (using simple random sampling 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 $\theta_1 < \theta_2$

Let $x_{(1)i}$ and $x_{(2)i}$ be the ith auxiliary variable at the first and second phase sampling respectively. y_2 be the study variable at the second phase sampling. Then

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

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

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

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

$$f(x) = \begin{cases} 1, & is\ the\ jth\ unit\ of\ population\ possessing\ ith\ auxiliary\ attributes.\ j = 1,2,\dots,n \\ 0, & otherwise \end{cases}$$

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

$$p_{(1)i} = \left(P_i + \overline{e}_{\tau(1)_i}\right);\quad p_{(2)i} = \left(P_i + \overline{e}_{\tau(2)_i}\right) \qquad (3)$$

for $E\left(\overline{e}_{\tau(1)_i}\right) = E\left(\overline{e}_{\tau(2)_i}\right) = 0$ (4)

and $C_y^2 = \frac{S_{y_2}^2}{\overline{Y}^2}$; $C_{\tau_1}^2 = \frac{S_{\tau_1}^2}{P^2}$; $\rho_{yx} = \frac{S_{yx}}{S_y S_x}$

2.2 Some other useful results

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

$$E(\overline{e}_{y_2})^2 = \theta_2 \overline{Y}^2 C_y^2 \quad ; \quad E(\overline{e}_{x(2)_i})^2 = \theta_2 \overline{X}_i^2 C_{x_i}^2 \quad ; \quad E(\overline{e}_{y_2}\overline{e}_{x(2)_i}) = \theta_2 \overline{YX}_i C_y C_{x_i}\rho_{yx_i}$$

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

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

$$E(\overline{e}_{y_2}\overline{e}_{x(1)_i}) = \theta_1 \overline{YX}_i C_y C_{x_i}\rho_{yx_i}$$

$$E\left(\overline{e}_{y_2}\left(\overline{e}_{x(1)_i} - \overline{e}_{x(2)_i}\right)\right) = (\theta_1 - \theta_2)\overline{YX}_i C_y C_{x_i}\rho_{yx_i}$$

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

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

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

$$E\left(\left(\overline{e}_{x(1)_i} - \overline{e}_{x(2)_i}\right)\left(\overline{e}_{x(1)_j} - \overline{e}_{x(2)_j}\right)\right) = (\theta_2 - \theta_1)\overline{X}_i \overline{X}_j C_{x_i} C_{x_j}\rho_{x_i x_j} \qquad for\ i \neq j$$

$$E(\overline{e}_{y_2}\overline{e}_{\tau(2)_i}) = \theta_2 \overline{Y} P_i C_y C_{\tau_i}\rho_{y\tau_i}$$

$$E\left(\overline{e}_{\tau(2)_i}\overline{e}_{\tau(2)_j}\right) = \theta_2 P_i P_j C_{\tau_i} C_{\tau_j}\rho_{\tau_j \tau_i} \qquad for\ i \neq j$$

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

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

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

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

According to [17]

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

2.3 Mixture Ratio Estimator in Two-phase Sampling.

2.3.1 Full Information Case (FIC)

The estimated population mean of a generalized mixture ratio estimator in two-phase sampling using multi-auxiliary attributes and variables when information on all the auxiliary attributes and variables are available from the population was established in [14]. This is called Full Information Case *(FIC)* and presented as$t_1 = \bar{y}_2 * \prod_{i=1}^{k}\left(\frac{\bar{X}_i}{\bar{x}_{(2)_i}}\right)^{\alpha_i}\prod_{j=k+1}^{q}\left(\frac{P_i}{p_{(2)_j}}\right)^{\beta_j}$ (5)

The corresponding Mean Square Error (MSE) is presented as

$$MSE(t_1) = \theta_2 \overline{Y}^2 C_y^2\left(1 - \rho_{y\left(\underset{\smile}{x},\underset{\smile}{\tau}\right)_q}^2\right)$$

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 in [14] as

$$t_2 = \bar{y}_2 * \prod_{i=1}^{k}\left(\frac{\bar{x}_{(1)_i}}{\bar{x}_{(2)_i}}\right)^{\alpha_i}\prod_{j=k+1}^{q}\left(\frac{p_{(1)_j}}{p_{(2)_j}}\right)^{\beta_j} \qquad (6)$$

The corresponding Mean Square Error (MSE) is given as

$$MSE(t_2) = \overline{Y}^2 C_y^2 \left[\theta_2 + (\theta_1 - \theta_2)\rho^2_{y.\underset{\sim}{x}_k} + (\theta_1 - \theta_2)\rho^2_{y.\underset{\sim}{\tau}_q}\right]$$

$$MSE(t_2) = \overline{Y}^2 C_y^2 \left[\theta_2\left(1 - \rho^2_{y\left(\underset{\sim}{x},\underset{\sim}{\tau}\right)_q}\right) + \theta_2 \rho^2_{y\left(\underset{\sim}{x},\underset{\sim}{\tau}\right)_q}\right]$$

2.3.3 Partial Information Case I (PIC-I)

The estimated population mean of mixture ratio estimator in two-phase sampling when there are multi-auxiliary variables and attributes was presented by [14]. This estimator assumes that there is no information on $(r+1),(r+2),\ldots,k$auxiliary variables and $(h+1),(h+2),\ldots,q$ auxiliary attributes from the population. This estimator is termed Partial Information Case I (PIC-I). This estimator uses the second method of configuring partial information case out of the two ways expressed by [7] of presenting partial information case. The estimator is presented as:

$$t_3 = \bar{y}_2\left[\prod_{i=1}^{r}\left(\frac{\bar{x}_{(1)_i}}{\bar{x}_{(2)_i}}\right)^{\alpha_i}\left(\frac{\bar{X}_i}{\bar{x}_{(1)_i}}\right)^{\beta_i}\right]\left[\prod_{j=r+1}^{k}\left(\frac{\bar{x}_{(1)_j}}{\bar{x}_{(2)_j}}\right)^{\alpha_j}\right]\left[\prod_{f=k+1}^{h}\left(\frac{p_{(1)_f}}{p_{(2)_f}}\right)^{\gamma_f}\left(\frac{P_f}{p_{(1)_f}}\right)^{\lambda_f}\right]\left[\prod_{g=h+1}^{q}\left(\frac{p_{(1)_g}}{p_{(2)_g}}\right)^{\gamma_g}\right] \qquad (7)$$

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

$$MSE(t_3) = \overline{Y}^2 C_y^2\left[\theta_2\left(1 - \rho^2_{y.\left(\underset{\sim}{x},\underset{\sim}{\tau}\right)q}\right) + \theta_1\left(\rho^2_{y.\left(\underset{\sim}{x},\underset{\sim}{\tau}\right)q} - \rho^2_{y.\left(\underset{\sim}{x},\underset{\sim}{\tau}\right)h}\right)\right]$$

3.0 Methodology

3.1 Introducing the Estimator Scheme

This research observes that in two-phase sampling, expressing some estimators on a paper page could span through the breadth of such page. The length of some estimators could be a threat to the understanding of such estimators. To understand the properties of and the difference between estimators start by placing such estimators side-by-side for comparison. The quality of this side-by-side estimators comparison may reduce if such estimators are lengthy. Considering this challenge, this research has proposed ESTIMATOR SCHEMA to represent an estimator which will reduce the length of any estimator.

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 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 of a ratio estimator in two-phase sampling is:

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

The Schema for the estimator t is presented as

$$t^* = \left[\overbrace{\bar{y}_2 * \underbrace{\alpha_i^{+r} * \beta_{1.i}^{+r}}_{\substack{AV\\PIC}} * \underbrace{\lambda_{2.j}^{+q}}_{\substack{AA\\FIC}}}^{Ratio\ (PIC)}\right] \begin{array}{l} \longrightarrow \textbf{LINE 1} \\ \longrightarrow \textbf{LINE 2} \\ \longrightarrow \textbf{LINE 3} \end{array}$$

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:

* $\alpha, \beta, and\ \lambda$ are parameters to be estimated in the estimator.
* $i\ and\ j$ are counters associated with the corresponding parameter. $i = 1,2,\ldots,r,\ \ j = r+1, r+2,\ldots,q.$
* $1.i$: FIC with the first phase sample data available about the auxiliary characteristic.
* $2.j$: FIC with the second phase sample data available about the auxiliary characteristic.

LINE 3: This is the type of auxiliary characteristics used. **AV** means Auxiliary Variable and **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*
Estimator schema, is hereby, introduced to the aforementioned estimators as proposed in [14] as thus:

a. Estimator Schema for Full Information Case (FIC)

The schema of estimator t_1 is presented as

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

b. Estimator Schema for No Information Case (NIC)

The schema of estimator t_2 is presented as

$$t_2^* = \left\{ \overbrace{\overline{y}_2 * \underbrace{\alpha_{i}^{+k}}_{\substack{AV\\NIC}} * \underbrace{\beta_{j}^{+q}}_{\substack{AA\\NIC}}}^{Ratio\ (NIC)} \right\}$$

c. Estimator Schema for Partial Information Case (PIC)

The schema of estimator t_3 is presented as

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

d. Estimator Schema Description

$\overline{y}_2 =$ Sample mean of the study variable at the second phase sampling

$\beta_{1.i}^{+r} = \left[\prod_{i=1}^{r}\left(\frac{\bar{X}_i}{\bar{x}_{(1)_i}}\right)^{\beta_i}\right]$: In t_3^* above, this is a full information case estimator with first phase sample data available. $i = 1,2,\ldots,r$. The **"+"** symbol before r means that the estimator uses ratio estimation method. However, the presence of **"-"** symbol means it is a product estimation method.

$\beta_{2.j}^{+q} = \left[\prod_{j=r+1}^{q}\left(\frac{\bar{X}_j}{\bar{x}_{(2)_j}}\right)^{\beta_j}\right]$: In t_1^* above, this is a full information case estimator with second phase sample data available. $j = (k+1),(k+2),\ldots,q$. The counter "j" initializes its counting from where the last counter (i) stopped its counting.

$\alpha_i^{+r} = \left[\prod_{i=1}^{g}\left(\frac{\bar{x}_{(1)_i}}{\bar{x}_{(2)_i}}\right)^{\alpha_i}\right]$: In t_3^* above, this is a no information case ratio estimator with $i = 1,2,\ldots,r$.

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. The auxiliary variable is at full information case while the auxiliary attribute is at partial information case. Then, the mixture ratio estimator is presented as:

$$t_4 = \overline{y}_2\left[\prod_{i=1}^{k}\left(\frac{\bar{x}_{(1)_i}}{\bar{x}_{(2)_i}}\right)^{\alpha_i}\left(\frac{\bar{X}_i}{\bar{x}_{(1)_i}}\right)^{\beta_i}\right]\left[\prod_{f=k+1}^{h}\left(\frac{p_{(1)_f}}{p_{(2)_f}}\right)^{\gamma_f}\left(\frac{P_f}{p_{(1)_f}}\right)^{\lambda_f}\right]\left[\prod_{g=h+1}^{q}\left(\frac{p_{(1)_g}}{p_{(2)_g}}\right)^{\gamma_g}\right] \quad (8)$$

The schema for estimator t_4 is presented as:

$$t_4^* = \left\{ \overbrace{\overline{y}_2 * \underbrace{\alpha_{i}^{+k} * \beta_{1.i}^{+k}}_{\substack{AV\\FIC}} * \underbrace{\gamma_f^{+h} * \lambda_{1.f}^{+h} * \gamma_g^{+q}}_{\substack{AA\\PIC}}}^{Ratio\ (PIC-II)} \right\} \quad (9)$$

Appling the equations (1) and (3) to equation (8) yields

$$t_4 = \left[(\bar{Y}+\bar{e}_{y2})\prod_{i=1}^{k}\left(\frac{\bar{X}_i+\bar{e}_{x(1)_i}}{\bar{X}_i+\bar{e}_{x(2)_i}}\right)^{\alpha_i}\left(\frac{\bar{X}_i}{\bar{X}_i+\bar{e}_{x(1)_i}}\right)^{\beta_i}\prod_{f=k+1}^{h}\left(\frac{P_f+\bar{e}_{\tau(1)_f}}{P_f+\bar{e}_{\tau(2)_f}}\right)^{\gamma_f}\left(\frac{P_f}{P_f+\bar{e}_{\tau(1)_f}}\right)^{\lambda_f}\prod_{g=h+1}^{q}\left(\frac{P_g+\bar{e}_{\tau(1)_g}}{P_g+\bar{e}_{\tau(2)_g}}\right)^{\gamma_g}\right]$$

$$t_4 = \left[\begin{array}{c}(\bar{Y}+\bar{e}_{y2})\prod_{i=1}^{k}\left(\frac{\bar{X}_i+\bar{e}_{x(1)_i}}{\bar{X}_i}\right)^{\alpha_i}\left(\frac{\bar{X}_i}{\bar{X}_i+\bar{e}_{x(2)_i}}\right)^{\alpha_i}\left(\frac{\bar{X}_i}{\bar{X}_i+\bar{e}_{x(1)_i}}\right)^{\beta_i} * \\ \prod_{f=k+1}^{h}\left(\frac{P_f+\bar{e}_{\tau(1)_f}}{P_f}\right)^{\gamma_f}\left(\frac{P_f}{P_f+\bar{e}_{\tau(2)_f}}\right)^{\gamma_f}\left(\frac{P_f}{P_f+\bar{e}_{\tau(1)_f}}\right)^{\lambda_f}\prod_{g=h+1}^{q}\left(\frac{P_g+\bar{e}_{\tau(1)_g}}{P_g}\right)^{\gamma_g}\left(\frac{P_g}{P_g+\bar{e}_{\tau(2)_g}}\right)^{\gamma_g}\end{array}\right]$$

$$t_4 = \left[\begin{array}{c}(\bar{Y}+\bar{e}_{y2})\prod_{i=1}^{k}\left(1+\frac{\bar{e}_{x(1)_i}}{\bar{X}_i}\right)^{\alpha_i}\left(1+\frac{\bar{e}_{x(2)_i}}{\bar{X}_i}\right)^{-\alpha_i}\left(1+\frac{\bar{e}_{x(1)_i}}{\bar{X}_i}\right)^{-\beta_i} * \\ \prod_{f=k+1}^{h}\left(1+\frac{\bar{e}_{\tau(1)_f}}{P_f}\right)^{\gamma_f}\left(1+\frac{\bar{e}_{\tau(2)_f}}{P_f}\right)^{-\gamma_f}\left(1+\frac{\bar{e}_{\tau(1)_f}}{P_f}\right)^{-\lambda_f}\prod_{g=h+1}^{q}\left(1+\frac{\bar{e}_{\tau(1)_g}}{P_g}\right)^{\gamma_g}\left(1+\frac{\bar{e}_{\tau(2)_g}}{P_g}\right)^{-\gamma_g}\end{array}\right]$$

Applying taylor's series of expansion, it gives thus:

$$t_4 = \left[(\bar{Y}+\bar{e}_{y2})\left(\begin{array}{c}\left(1+\sum_{i=1}^{k}\alpha_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i}\right)\left(1-\sum_{i=1}^{k}\alpha_i\frac{\bar{e}_{x(2)_i}}{\bar{X}_i}\right)\sum_{i=1}^{k}\left(1-\sum_{i=1}^{k}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i}\right)\left(1+\sum_{f=k+1}^{h}\gamma_f\frac{\bar{e}_{\tau(1)_f}}{P_f}\right)\left(1-\sum_{f=k+1}^{h}\gamma_f\frac{\bar{e}_{\tau(2)_f}}{P_f}\right) \\ \left(1-\sum_{f=k+1}^{h}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f}\right)\left(1+\sum_{g=h+1}^{q}\gamma_g\frac{\bar{e}_{\tau(1)_g}}{P_g}\right)\left(1-\sum_{g=h+1}^{q}\gamma_g\frac{\bar{e}_{\tau(2)_g}}{P_g}\right)\end{array}\right)\right]$$

Expand the brackets ignoring the second and higher order terms to give:

$$t_4 = \left[(\bar{Y}+\bar{e}_{y2})\left(\begin{array}{c}1+\sum_{i=1}^{k}\alpha_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i}-\sum_{i=1}^{k}\alpha_i\frac{\bar{e}_{x(2)_i}}{\bar{X}_i}-\sum_{i=1}^{k}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i}+\sum_{f=k+1}^{h}\gamma_f\frac{\bar{e}_{\tau(1)_f}}{P_f}-\sum_{f=k+1}^{h}\gamma_f\frac{\bar{e}_{\tau(2)_f}}{P_f}-\sum_{f=k+1}^{h}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f}+ \\ \sum_{g=h+1}^{q}\gamma_g\frac{\bar{e}_{\tau(1)_g}}{P_g}-\sum_{g=h+1}^{q}\gamma_g\frac{\bar{e}_{\tau(2)_g}}{P_g}\end{array}\right)\right]$$

Expand the brackets ignoring the second and higher order terms to give:

$$t_4 = \left[\bar{Y}+\bar{e}_{y2}+\bar{Y}\sum_{i=1}^{k}\alpha_i\frac{\left(\bar{e}_{x(1)_i}-\bar{e}_{x(2)_i}\right)}{\bar{X}_i}-\bar{Y}\sum_{i=1}^{k}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i}+\bar{Y}\sum_{f=k+1}^{h}\gamma_f\frac{\left(\bar{e}_{\tau(1)_f}-\bar{e}_{\tau(2)_f}\right)}{P_f}-\bar{Y}\sum_{f=k+1}^{h}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f}\right.$$
$$\left.+\bar{Y}\sum_{g=h+1}^{q}\gamma_g\frac{\left(\bar{e}_{\tau(1)_g}-\bar{e}_{\tau(2)_g}\right)}{P_g}\right]$$

$$MSE(t_4) = E_1E_{2/1}\left(t_4-\bar{Y}\right)^2$$

$$MSE(t_4) = E_1E_{2/1}\left[\bar{e}_{y2}+\bar{Y}\sum_{i=1}^{k}\alpha_i\frac{\left(\bar{e}_{x(1)_i}-\bar{e}_{x(2)_i}\right)}{\bar{X}_i}-\bar{Y}\sum_{i=1}^{k}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i}+\bar{Y}\sum_{f=k+1}^{h}\gamma_f\frac{\left(\bar{e}_{\tau(1)_f}-\bar{e}_{\tau(2)_f}\right)}{P_f}\right.$$
$$-\bar{Y}\sum_{f=k+1}^{h}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f}$$
$$\left.+\bar{Y}\sum_{g=h+1}^{q}\gamma_g\frac{\left(\bar{e}_{\tau(1)_g}-\bar{e}_{\tau(2)_g}\right)}{P_g}\right]^2 \qquad (10)$$

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 equate it to zero, hence, solve for the parameters.

$$\frac{\partial MSE(t_4)}{\partial \alpha_i} = 0 \qquad\qquad for\ i = 1,2,\dots,k$$

$$E_1E_{2/1}\left[\bar{Y}\frac{\left(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i}\right)}{\bar{X}_i}\left(\bar{e}_{y2} + \bar{Y}\sum_{i=1}^{k}\alpha_i\frac{\left(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i}\right)}{\bar{X}_i} - \bar{Y}\sum_{i=1}^{k}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i} + \bar{Y}\sum_{f=k+1}^{h}\gamma_f\frac{\left(\bar{e}_{\tau(1)_f} - \bar{e}_{\tau(2)_f}\right)}{P_f} - \bar{Y}\sum_{f=k+1}^{h}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f} + \bar{Y}\sum_{g=h+1}^{q}\gamma_g\frac{\left(\bar{e}_{\tau(1)_g} - \bar{e}_{\tau(2)_g}\right)}{P_g}\right)\right] = 0$$

Applying expectation and simplifying further when $i = 1$

$$\sum_{i=1}^{k}\alpha_iC_{x_i}\rho_{x_1x_i} + \sum_{f=k+1}^{h}\gamma_fC_{\tau_f}\rho_{x_1\tau_f} + \sum_{g=h+1}^{q}\gamma_gC_{\tau_g}\rho_{x_1\tau_g} = C_y\rho_{yx_1} \qquad (11)$$

Applying expectation and simplifying further when $i = 2$

$$\sum_{i=1}^{k}\alpha_iC_{x_i}\rho_{x_2x_i} + \sum_{f=k+1}^{h}\gamma_fC_{\tau_f}\rho_{x_2\tau_f} + \sum_{g=h+1}^{q}\gamma_gC_{\tau_g}\rho_{x_2\tau_g} = C_y\rho_{yx_2} \qquad (12)$$

$$\dots\ \dots\ \dots\ \dots.\ \dots\ \dots\ \dots\ \dots.\ \dots\ \dots\ \dots\ \dots.$$

Applying expectation and simplifying further when $i = k$

$$\sum_{i=1}^{k}\alpha_iC_{x_i}\rho_{x_kx_i} + \sum_{f=k+1}^{h}\gamma_fC_{\tau_f}\rho_{x_k\tau_f} + \sum_{g=h+1}^{q}\gamma_gC_{\tau_g}\rho_{x_k\tau_g} = C_y\rho_{yx_k} \qquad (13)$$

$$\frac{\partial MSE(t_4)}{\partial \beta_i} = 0 \qquad\qquad for\ i = 1,2,\dots,k$$

$$E_1E_{2/1}\left[-\bar{Y}\frac{\bar{e}_{x(1)_i}}{\bar{X}_i}\left(\bar{e}_{y2} + \bar{Y}\sum_{i=1}^{k}\alpha_i\frac{\left(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i}\right)}{\bar{X}_i} - \bar{Y}\sum_{i=1}^{k}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i} + \bar{Y}\sum_{f=k+1}^{h}\gamma_f\frac{\left(\bar{e}_{\tau(1)_f} - \bar{e}_{\tau(2)_f}\right)}{P_f} - \bar{Y}\sum_{f=k+1}^{h}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f} + \bar{Y}\sum_{g=h+1}^{q}\gamma_g\frac{\left(\bar{e}_{\tau(1)_g} - \bar{e}_{\tau(2)_g}\right)}{P_g}\right)\right] = 0$$

Applying expectation and simplifying further when $i = 1$

$$\sum_{i=1}^{k}\beta_iC_{x_i}\rho_{x_1x_i} + \sum_{f=k+1}^{h}\lambda_fC_{\tau_f}\rho_{x_1\tau_f} = C_y\rho_{yx_1} \qquad (14)$$

Applying expectation and simplifying further when $i = 2$

$$\sum_{i=1}^{k}\beta_iC_{x_i}\rho_{x_2x_i} + \sum_{f=k+1}^{h}\lambda_fC_{\tau_f}\rho_{x_2\tau_f} = C_y\rho_{yx_2} \qquad (15)$$

$$\dots\ \dots\ \dots\ \dots.\ \dots\ \dots\ \dots\ \dots.\ \dots\ \dots\ \dots\ \dots.$$

Applying expectation and simplifying further when $i = k$

$$\sum_{i=1}^{k}\beta_iC_{x_i}\rho_{x_kx_i} + \sum_{f=k+1}^{h}\lambda_fC_{\tau_f}\rho_{x_k\tau_f} = C_y\rho_{yx_k} \qquad (16)$$

$$\frac{\partial MSE(t_4)}{\partial \gamma_f} = 0 \qquad\qquad for\ f = k+1, k+2,\dots,h$$

$$E_1E_{2/1}\left[\bar{Y}\frac{\left(\bar{e}_{\tau(1)_f}-\bar{e}_{\tau(2)_f}\right)}{P_f}\left(\bar{e}_{y2}+\bar{Y}\sum_{i=1}^{k}\alpha_i\frac{\left(\bar{e}_{x(1)_i}-\bar{e}_{x(2)_i}\right)}{\bar{X}_i}-\bar{Y}\sum_{i=1}^{k}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i}+\bar{Y}\sum_{f=k+1}^{h}\gamma_f\frac{\left(\bar{e}_{\tau(1)_f}-\bar{e}_{\tau(2)_f}\right)}{P_f}-\bar{Y}\sum_{f=k+1}^{h}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f}+\bar{Y}\sum_{g=h+1}^{q}\gamma_g\frac{\left(\bar{e}_{\tau(1)_g}-\bar{e}_{\tau(2)_g}\right)}{P_g}\right)\right]=0$$

Applying expectation and simplifying further when $f=(k+1)$

$$\sum_{i=1}^{k}\alpha_iC_{x_i}\rho_{\tau_{k+1}x_i}+\sum_{f=k+1}^{h}\gamma_fC_{\tau_f}\rho_{\tau_{k+1}\tau_f}+\sum_{g=h+1}^{q}\gamma_gC_{\tau_g}\rho_{\tau_{k+1}\tau_g}=C_y\rho_{y\tau_{k+1}} \qquad (17)$$

Applying expectation and simplifying further when $f=(k+2)$

$$\sum_{i=1}^{k}\alpha_iC_{x_i}\rho_{\tau_{k+2}x_i}+\sum_{f=k+1}^{h}\gamma_fC_{\tau_f}\rho_{\tau_{k+2}\tau_f}+\sum_{g=h+1}^{q}\gamma_gC_{\tau_g}\rho_{\tau_{k+2}\tau_g}=C_y\rho_{y\tau_{k+2}} \qquad (18)$$

… … … …. … … … …. … … … ….

Applying expectation and simplifying further when $f=h$

$$\sum_{i=1}^{k}\alpha_iC_{x_i}\rho_{\tau_hx_i}+\sum_{f=k+1}^{h}\gamma_fC_{\tau_f}\rho_{\tau_h\tau_f}+\sum_{g=h+1}^{q}\gamma_gC_{\tau_g}\rho_{\tau_h\tau_g}=C_y\rho_{y\tau_h} \qquad (19)$$

$$\frac{\partial MSE(t_4)}{\partial\lambda_f}=0 \qquad for\ f=k+1,k+2,\dots,h$$

$$E_1E_{2/1}\left[-\bar{Y}\frac{\bar{e}_{\tau(1)_f}}{P_f}\left(\bar{e}_{y2}+\bar{Y}\sum_{i=1}^{k}\alpha_i\frac{\left(\bar{e}_{x(1)_i}-\bar{e}_{x(2)_i}\right)}{\bar{X}_i}-\bar{Y}\sum_{i=1}^{k}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i}+\bar{Y}\sum_{f=k+1}^{h}\gamma_f\frac{\left(\bar{e}_{\tau(1)_f}-\bar{e}_{\tau(2)_f}\right)}{P_f}-\bar{Y}\sum_{f=k+1}^{h}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f}+\bar{Y}\sum_{g=h+1}^{q}\gamma_g\frac{\left(\bar{e}_{\tau(1)_g}-\bar{e}_{\tau(2)_g}\right)}{P_g}\right)\right]=0$$

Applying expectation and simplifying further when $f=k+1$

$$\sum_{i=1}^{k}\beta_iC_{x_i}\rho_{\tau_{k+1}x_i}+\sum_{f=k+1}^{h}\lambda_fC_{\tau_f}\rho_{\tau_{k+1}\tau_f}=C_y\rho_{y\tau_{k+1}} \qquad (20)$$

Applying expectation and simplifying further when $f=k+2$

$$\sum_{i=1}^{k}\beta_iC_{x_i}\rho_{\tau_{k+2}x_i}+\sum_{f=k+1}^{h}\lambda_fC_{\tau_f}\rho_{\tau_{k+2}\tau_f}=C_y\rho_{y\tau_{k+2}} \qquad (21)$$

… … … …. … … … …. … … … ….

Applying expectation and simplifying further when $f=h$

$$\sum_{i=1}^{k}\beta_iC_{x_i}\rho_{\tau_hx_i}+\sum_{f=k+1}^{h}\lambda_fC_{\tau_f}\rho_{\tau_h\tau_f}=C_y\rho_{y\tau_h} \qquad (22)$$

$$\frac{\partial MSE(t_4)}{\partial\gamma_g}=0 \qquad for\ g=h+1,h+2,\dots,q$$

$$E_1E_{2/1}\left[\bar{Y}\frac{\left(\bar{e}_{\tau(1)_g}-\bar{e}_{\tau(2)_g}\right)}{P_g}\left(\bar{e}_{y2}+\bar{Y}\sum_{i=1}^{k}\alpha_i\frac{\left(\bar{e}_{x(1)_i}-\bar{e}_{x(2)_i}\right)}{\bar{X}_i}-\bar{Y}\sum_{i=1}^{k}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i}+\bar{Y}\sum_{f=k+1}^{h}\gamma_f\frac{\left(\bar{e}_{\tau(1)_f}-\bar{e}_{\tau(2)_f}\right)}{P_f}-\bar{Y}\sum_{f=k+1}^{h}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f}+\bar{Y}\sum_{g=h+1}^{q}\gamma_g\frac{\left(\bar{e}_{\tau(1)_g}-\bar{e}_{\tau(2)_g}\right)}{P_g}\right)\right]=0$$

Applying expectation and simplifying further when $g = (h+1)$

$$\sum_{i=1}^{k} \alpha_i C_{x_i} \rho_{\tau_{h+1} x_i} + \sum_{f=k+1}^{h} \gamma_f C_{\tau_f} \rho_{\tau_{h+1} \tau_f} + \sum_{g=h+1}^{q} \gamma_g C_{\tau_g} \rho_{\tau_{h+1} \tau_g} = C_y \rho_{y \tau_{h+1}} \quad (23)$$

Applying expectation and simplifying further when $g = (h+2)$

$$\sum_{i=1}^{k} \alpha_i C_{x_i} \rho_{\tau_{h+2} x_i} + \sum_{f=k+1}^{h} \gamma_f C_{\tau_f} \rho_{\tau_{h+2} \tau_f} + \sum_{g=h+1}^{q} \gamma_g C_{\tau_g} \rho_{\tau_{h+2} \tau_g} = C_y \rho_{y \tau_{h+2}} \quad (24)$$

… … … …. … … … …. … … … ….

Applying expectation and simplifying further when $g = q$

$$\sum_{i=1}^{k} \alpha_i C_{x_i} \rho_{\tau_q x_i} + \sum_{f=k+1}^{h} \gamma_f C_{\tau_f} \rho_{\tau_q \tau_f} + \sum_{g=h+1}^{q} \gamma_g C_{\tau_g} \rho_{\tau_q \tau_g} = C_y \rho_{y \tau_q} \quad (25)$$

Representing equations (11),(12), (13), (17), (18), (19), (23), (24) and (25) in correlation coefficient matrix and simplifying further to give:

$$\sigma_{*k} = \frac{C_y (-1)^{i+1} \left| R_{y x_i \tau_j} \right|_{y(\underset{\smile}{x}, \underset{\smile}{\tau})_k}}{|R|_{y(\underset{\smile}{x}, \underset{\smile}{\tau})_k}} = C_{*k} \sigma^{*k}$$

Hence, the following parameters are estimated as thus:

$$\alpha_i = \frac{C_y (-1)^{i+1} \left| R_{y x_i} \right|_{y \underset{\smile}{x}_k}}{C_{x_i} |R|_{\underset{\smile}{x}_k}} \qquad for\ i = 1,2, \dots, k \quad (26)$$

$$\gamma_f = \frac{C_y (-1)^{f+1} \left| R_{y \tau_f} \right|_{y \underset{\smile}{\tau}_h}}{C_{\tau_f} |R|_{\underset{\smile}{\tau}_h}} \qquad for\ f = k+1, k+2, \dots, h \quad (27)$$

$$\gamma_g = \frac{C_y (-1)^{g+1} \left| R_{y \tau_g} \right|_{y \underset{\smile}{\tau}_q}}{C_{\tau_g} |R|_{\underset{\smile}{\tau}_q}} \qquad for\ g = f+1, f+2, \dots, q \quad (28)$$

Representing equations (14), (15), (16), (20), (21) and (22) in correlation coefficient matrix and simplifying further to give:

$$\sigma_{*h} = \frac{C_y (-1)^{i+1} \left| R_{y x_i, \tau_j} \right|_{y(\underset{\smile}{x}, \underset{\smile}{\tau})_h}}{|R|_{y(\underset{\smile}{x}, \underset{\smile}{\tau})_h}} = C_{*h} \sigma^{*h}$$

Hence, the following parameters are estimated as thus:

$$\beta_i = \frac{C_y (-1)^{i+1} \left| R_{y x_i} \right|_{y \underset{\smile}{x}_k}}{C_{x_i} |R|_{\underset{\smile}{x}_k}} \qquad for\ i = 1, 2, \dots, k \quad (29)$$

$$\lambda_f = \frac{C_y (-1)^{f+1} \left| R_{y \tau_f} \right|_{y \underset{\smile}{\tau}_h}}{C_{\tau_f} |R|_{\underset{\smile}{\tau}_h}} \qquad for\ f = k+1, k+2, \dots, h \quad (30)$$

Simplifying equation (10) ignoring the second and higher order degrees to give

$$MSE(t_4) = E_1E_{2/1}\left[\bar{e}_{y2}\left(\bar{e}_{y2} + \bar{Y}\sum_{i=1}^{k}\alpha_i\frac{\left(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i}\right)}{\bar{X}_i} - \bar{Y}\sum_{i=1}^{k}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i} + \bar{Y}\sum_{f=k+1}^{h}\gamma_f\frac{\left(\bar{e}_{\tau(1)_f} - \bar{e}_{\tau(2)_f}\right)}{P_f}\right.\right.$$
$$- \bar{Y}\sum_{f=k+1}^{h}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f}$$
$$\left.\left. + \bar{Y}\sum_{g=h+1}^{q}\gamma_g\frac{\left(\bar{e}_{\tau(1)_g} - \bar{e}_{\tau(2)_g}\right)}{P_g}\right)\right] \quad (31)$$

Applying conditional expectation to equation (31) gives

$$MSE(t_4) = \bar{Y}^2C_y\left[\theta_2C_y + (\theta_1 - \theta_2)\sum_{i=1}^{k}\alpha_iC_{x_i}\rho_{yx_i} - \theta_1\sum_{i=1}^{k}\beta_iC_{x_i}\rho_{yx_i} + (\theta_1 - \theta_2)\sum_{f=k+1}^{h}\gamma_fC_{\tau_f}\rho_{y\tau_f}\right.$$
$$\left. - \theta_1\sum_{f=k+1}^{h}\lambda_fC_{\tau_f}\rho_{y\tau_f} + (\theta_1 - \theta_2)\sum_{g=h+1}^{q}\gamma_gC_{\tau_g}\rho_{y\tau_g}\right] \quad (32)$$

Substitute the optimum equations obtained for α_i , β_i, γ_f ,λ_f and γ_g, in equations (26), (27), (28), (29) and (30) into equation (32), hence, simplify to give:

$$MSE(t_4)_{min} = \bar{Y}^2C_y^2\left[\theta_2 + (\theta_1 - \theta_2)\rho^2_{y.\underset{\smile}{x}_k} - \theta_1\rho^2_{y.\underset{\smile}{x}_k} + (\theta_1 - \theta_2)\rho^2_{y.\underset{\smile}{\tau}_h} - \theta_1\rho^2_{y.\underset{\smile}{\tau}_h} + (\theta_1 - \theta_2)\rho^2_{y.\underset{\smile}{\tau}_q}\right]$$
$$MSE(t_4)_{min} = \bar{Y}^2C_y^2\left[\theta_2 + (\theta_1 - \theta_2)\rho^2_{y.\left(\underset{\smile}{x},\underset{\smile}{\tau}\right)_q} - \theta_1\rho^2_{y.\left(\underset{\smile}{x}_k,\underset{\smile}{\tau}_h\right)}\right]$$
$$MSE(t_4)_{min} = \bar{Y}^2C_y^2\left[\theta_2\left(1 - \rho^2_{y.\left(\underset{\smile}{x},\underset{\smile}{\tau}\right)_q}\right) + \theta_1\left(\rho^2_{y.\left(\underset{\smile}{x},\underset{\smile}{\tau}\right)_q} - \rho^2_{y.\left(\underset{\smile}{x}_k,\underset{\smile}{\tau}_h\right)}\right)\right] \quad (33)$$

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 $1\ to\ r$ 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. The auxiliary variable is at partial information case while the auxiliary attribute is at full information case. The estimator is suggested as:

$$t_5 = \bar{y}_2 * \left[\prod_{i=1}^{r}\left(\frac{\bar{x}_{(1)_i}}{\bar{x}_{(2)_i}}\right)^{\alpha_i}\left(\frac{\bar{X}_i}{\bar{x}_{(1)_i}}\right)^{\beta_i}\right]\left[\prod_{j=r+1}^{k}\left(\frac{\bar{x}_{(1)_j}}{\bar{x}_{(2)_j}}\right)^{\alpha_j}\right]\left[\prod_{f=k+1}^{q}\left(\frac{p_{(1)_f}}{p_{(2)_f}}\right)^{\gamma_f}\left(\frac{P_f}{p_{(1)_f}}\right)^{\lambda_f}\right] \quad (34)$$

The schema for estimator t_5 is presented as:

$$t_5^* = \left\{\overbrace{\bar{y}_2 * \underbrace{\alpha_i^{+r} * \beta_{1.i}^{+r} * \alpha_j^{+k}}_{\substack{AV\\PIC}} * \underbrace{\gamma_f^{+q} * \lambda_{1.f}^{+q}}_{\substack{AA\\FIC}}}^{Ratio\ (PIC-II)}\right\} \quad (35)$$

The proof for the mean square error of estimator t_5 is abridged. However, it follows the same steps as established in PIC-II explained earlier. Applying equation (1) and (3) to equation (34) gives

$$MSE(t_5) = E_1E_{2/1}\left[\bar{e}_{y2} + \bar{Y}\sum_{i=1}^{r}\alpha_i\frac{\left(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i}\right)}{\bar{X}_i} - \bar{Y}\sum_{i=1}^{r}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i} + \bar{Y}\sum_{j=r+1}^{k}\alpha_j\frac{\left(\bar{e}_{x(1)_j} - \bar{e}_{x(2)_j}\right)}{\bar{X}_j}\right.$$
$$\left. + \bar{Y}\sum_{f=k+1}^{q}\gamma_f\frac{\left(\bar{e}_{\tau(1)_f} - \bar{e}_{\tau(2)_f}\right)}{P_f} - \bar{Y}\sum_{f=k+1}^{q}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f}\right]^2 \quad (36)$$

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

$$\alpha_i = \frac{C_y(-1)^{i+1}\left|R_{yx_i}\right|_{y\underset{\smile}{x}_r}}{C_{x_i}|R|_{\underset{\smile}{x}_r}} \qquad for\ i = 1,2,\dots,r \qquad (37)$$

$$\alpha_j = \frac{C_y(-1)^{j+1}\left|R_{yx_j}\right|_{y\underset{\smile}{x}_k}}{C_{x_j}|R|_{\underset{\smile}{x}_k}} \qquad for\ j = r+1, r+2,\dots,k \qquad (38)$$

$$\gamma_f = \frac{C_y(-1)^{f+1}\left|R_{y\tau_f}\right|_{y\underset{\smile}{\tau}_q}}{C_{\tau_f}|R|_{\underset{\smile}{\tau}_q}} \qquad for\ f = k+1, k+2,\dots,q \qquad (39)$$

$$\beta_i = \frac{C_y(-1)^{i+1}\left|R_{yx_i}\right|_{y\underset{\smile}{x}_r}}{C_{x_i}|R|_{\underset{\smile}{x}_r}} \qquad for\ i = 1,2,\dots,r \qquad (40)$$

$$\lambda_f = \frac{C_y(-1)^{f+1}\left|R_{y\tau_f}\right|_{y\underset{\smile}{\tau}_q}}{C_{\tau_f}|R|_{\underset{\smile}{\tau}_q}} \qquad for\ f = k+1, k+2,\dots,q \qquad (41)$$

Simplify equation (36) and ignoring second and higher order degrees to give:

$$MSE(t_5) = E_1E_{2/1}\left[\bar{e}_{y2}\left(\bar{e}_{y2} + \bar{Y}\sum_{i=1}^{r}\alpha_i\frac{\left(\bar{e}_{x(1)_i} - \bar{e}_{x(2)_i}\right)}{\bar{X}_i} - \bar{Y}\sum_{i=1}^{r}\beta_i\frac{\bar{e}_{x(1)_i}}{\bar{X}_i} + \bar{Y}\sum_{j=r+1}^{k}\alpha_j\frac{\left(\bar{e}_{x(1)_j} - \bar{e}_{x(2)_j}\right)}{\bar{X}_j} + \bar{Y}\sum_{f=k+1}^{q}\gamma_f\frac{\left(\bar{e}_{\tau(1)_f} - \bar{e}_{\tau(2)_f}\right)}{P_f} - \bar{Y}\sum_{f=k+1}^{q}\lambda_f\frac{\bar{e}_{\tau(1)_f}}{P_f}\right)\right] \qquad (42)$$

Applying expectation to equation (42) and substitute the optimum equations obtained for α_i , β_i, α_j , γ_f $and\ \lambda_f$ in equations (37), (38), (39), (40) and (41) respectively into equation (42) then simplify further to give:

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

$$MSE(t_5)_{min} = \bar{Y}^2C_y^2\left[\theta_2 + (\theta_1-\theta_2)\rho^2_{y.\left(\underset{\smile}{x},\underset{\smile}{\tau}\right)_q} - \theta_1\rho^2_{y.\left(\underset{\smile}{x}_r,\underset{\smile}{\tau}_q\right)}\right]$$

$$MSE(t_5)_{min} = \bar{Y}^2C_y^2\left[\theta_2\left(1 - \rho^2_{y.\left(\underset{\smile}{x},\underset{\smile}{\tau}\right)_q}\right) + \theta_1\left(\rho^2_{y.\left(\underset{\smile}{x},\underset{\smile}{\tau}\right)_q} - \rho^2_{y.\left(\underset{\smile}{x}_r,\underset{\smile}{\tau}_q\right)}\right)\right] \qquad (43)$$

4.0 Comparison, Discussion and Conclusion

4.1 Comparison of the Estimators in PIC-I, PIC-II, PIC-III and NIC.

This section will perform theoretical and empirical comparison of the proposed estimators with the reviewed estimators.

4.1.1 *Theoretical Comparison of PIC-I and PIC-II*

$MSE(t_3)_{min} - MSE(t_4)_{min}$

$$\bar{Y}^2C_y^2\left[\theta_2\left(1 - \rho^2_{y.\left(\underset{\smile}{x},\underset{\smile}{\tau}\right)_q}\right) + \theta_1\left(\rho^2_{y.\left(\underset{\smile}{x},\underset{\smile}{\tau}\right)_q} - \rho^2_{y.\left(\underset{\smile}{x},\underset{\smile}{\tau}\right)_h}\right)\right] - \bar{Y}^2C_y^2\left[\theta_2\left(1 - \rho^2_{y.\left(\underset{\smile}{x},\underset{\smile}{\tau}\right)_q}\right) + \theta_1\left(\rho^2_{y.\left(\underset{\smile}{x},\underset{\smile}{\tau}\right)_q} - \rho^2_{y.\left(\underset{\smile}{x}_k,\underset{\smile}{\tau}_h\right)}\right)\right]$$

$$-\rho^2_{y.\left(\underset{\smile}{x},\underset{\smile}{\tau}\right)_h} + \rho^2_{y.\left(\underset{\smile}{x}_k,\underset{\smile}{\tau}_h\right)} > 0 \qquad (44)$$

For $k > h$ then $(MSE(t_3)_{min} > MSE(t_4)_{min})$. Hence, t_4 will be efficient than t_3 $(since\ k > h)$.

4.1.2 *Theoretical Comparison of PIC-I and PIC-III*

$$MSE(t_3)_{min} - MSE(t_5)_{min}$$

$$\bar{Y}^2 C_y^2 \left[\theta_2 \left(1 - \rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q}\right) + \theta_1 \left(\rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q} - \rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_h}\right)\right] - \bar{Y}^2 C_y^2 \left[\theta_2 \left(1 - \rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q}\right) + \theta_1 \left(\rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q} - \rho^2_{y.(\underset{\sim}{x}_r,\underset{\sim}{\tau}_q)}\right)\right]$$

$$-\rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_h} + \rho^2_{y.(\underset{\sim}{x}_r,\underset{\sim}{\tau}_q)} > 0 \qquad (45)$$

For $q > h$ and $> r$, then either $r > h$ or $h > r$ then, t_5 will be efficient than t_3.

4.1.3 *Theoretical Comparison of PIC-II and PIC-III*

$$MSE(t_4)_{min} - MSE(t_5)_{min}$$

$$\bar{Y}^2 C_y^2 \left[\theta_2 \left(1 - \rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q}\right) + \theta_1 \left(\rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q} - \rho^2_{y.(\underset{\sim}{x}_k,\underset{\sim}{\tau}_h)}\right)\right] - \bar{Y}^2 C_y^2 \left[\theta_2 \left(1 - \rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q}\right) + \theta_1 \left(\rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q} - \rho^2_{y.(\underset{\sim}{x}_r,\underset{\sim}{\tau}_q)}\right)\right]$$

$$-\rho^2_{y.(\underset{\sim}{x}_k,\underset{\sim}{\tau}_h)} + \rho^2_{y.(\underset{\sim}{x}_r,\underset{\sim}{\tau}_q)} \qquad (46)$$

For $q > h$, $q = k$ and $k > r$.
t_5 will be efficient than t_4 if and only if $r > h$ else t_4 will be efficient than t_5.

4.1.4 *Theoretical Comparison of PIC-II and NIC*

$$MSE(t_4)_{min} - MSE(t_2)_{min}$$

$$\bar{Y}^2 C_y^2 \left[\theta_2 \left(1 - \rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q}\right) + \theta_1 \left(\rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q} - \rho^2_{y.(\underset{\sim}{x}_k,\underset{\sim}{\tau}_h)}\right)\right] - \bar{Y}^2 C_y^2 \left[\theta_2 \left(1 - \rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q}\right) + \theta_1 \rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q}\right]$$

$$-\rho^2_{y.(\underset{\sim}{x}_k,\underset{\sim}{\tau}_h)} < 0 \qquad (47)$$

Since $\left(-\rho^2_{y.(\underset{\sim}{x}_k,\underset{\sim}{\tau}_h)} < 0\right)$ then t_4 is efficient than t_2.

4.1.5 *Comparison of PIC-III and NIC*

$$MSE(t_5)_{min} - MSE(t_2)_{min}$$

$$\bar{Y}^2 C_y^2 \left[\theta_2 \left(1 - \rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q}\right) + \theta_1 \left(\rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q} - \rho^2_{y.(\underset{\sim}{x}_r,\underset{\sim}{\tau}_q)}\right)\right] - \bar{Y}^2 C_y^2 \left[\theta_2 \left(1 - \rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q}\right) + \theta_1 \rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_q}\right]$$

$$-\rho^2_{y.(\underset{\sim}{x}_r,\underset{\sim}{\tau}_q)} < 0 \qquad (48)$$

Since $-\rho^2_{y.(\underset{\sim}{x}_r,\underset{\sim}{\tau}_q)} < 0$ then t_5 is efficient than t_2.

4.1.6 *Empirical Comparison of all Estimators*

This research uses *R* statistical software for simulation and analysis of the empirical comparison. Table 1 shows the corresponding population size, first phase and second phase sample sizes used in the analysis. The multiple correlation coefficient used in the analysis are presented in table 1. This research uses estimator average rating method of [18] as shown in table 1. Table 2 shows the ranking of all the five information cases in sixteen (16) population cases after which the overall average ranking is done. The analysis revealed that Full Information Case (FIC) claims to be the most efficient estimator and No Information Case (NIC) claims to be the least efficient estimator.

It is also revealed that Partial Information Case III (PIC-III) is efficient over Partial Information Case II (PIC-II) and Partial Information Case I (PIC-I). PIC-II proves efficient over PIC-I. This shows that the proposed estimators (PIC-II and PIC-III) are efficient than the PIC-I established in [14]. The R code used in this empirical analysis is available as free-and-open source code on github.com *at https://github.com/ogunyinka/mr/tree/mrr* .

Table 1: Means, Coefficient of Variation and Correlation Coefficient.

Pop.	N	n_1	n_2	***Seed***	$\overline{Y}^2$	C_y^2	$\rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_5}$	$\rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_3}$	$\rho^2_{y.(\underset{\sim}{x},\underset{\sim}{\tau})_{4-5}}$	$\rho^2_{y.(\underset{\sim}{x}_5,\underset{\sim}{\tau}_3)}$	$\rho^2_{y.(\underset{\sim}{x}_3,\underset{\sim}{\tau}_5)}$
1	*10000*	*3333*	*1111*	*800*	*710813.93*	*0.2969*	*0.9574*	*0.9535*	*0.955*	*0.9536*	*0.9573*
2	*9350*	*3117*	*1039*	*215*	*618230.63*	*0.2860*	*0.9552*	*0.9515*	*0.953*	*0.9517*	*0.9550*
3	*8700*	*2900*	*967*	*818*	*563605.14*	*0.2867*	*0.9577*	*0.9550*	*0.956*	*0.9557*	*0.9570*
4	*8050*	*2683*	*894*	*506*	*486608.18*	*0.2841*	*0.9576*	*0.9535*	*0.953*	*0.9536*	*0.9575*
5	*7400*	*2467*	*822*	*286*	*439636.09*	*0.2812*	*0.9593*	*0.9564*	*0.955*	*0.9564*	*0.9592*
6	*6750*	*2250*	*750*	*569*	*368315.85*	*0.2723*	*0.9588*	*0.9562*	*0.955*	*0.9563*	*0.9587*
7	*6100*	*2033*	*678*	*569*	*312509.93*	*0.2608*	*0.9583*	*0.9554*	*0.953*	*0.9555*	*0.9583*
8	*5450*	*1817*	*606*	*250*	*267927.20*	*0.2417*	*0.9584*	*0.9543*	*0.952*	*0.9543*	*0.9584*
9	*4800*	*1600*	*533*	*81*	*222415.93*	*0.2285*	*0.9588*	*0.9535*	*0.952*	*0.9537*	*0.9586*
10	*4150*	*1383*	*461*	*374*	*177689.51*	*0.2121*	*0.9578*	*0.9521*	*0.950*	*0.9524*	*0.9572*
11	*3500*	*1167*	*389*	*24*	*138773.92*	*0.1804*	*0.9544*	*0.9513*	*0.942*	*0.9514*	*0.9542*
12	*2850*	*950*	*317*	*101*	*109591.37*	*0.1779*	*0.9611*	*0.9585*	*0.948*	*0.9587*	*0.9609*
13	*2200*	*733*	*244*	*259*	*79513.43*	*0.1411*	*0.9591*	*0.9537*	*0.932*	*0.9549*	*0.9578*
14	*1550*	*517*	*172*	*598*	*58405.81*	*0.1193*	*0.9663*	*0.9543*	*0.910*	*0.9548*	*0.9658*
15	*900*	*300*	*100*	*357*	*35949.22*	*0.0641*	*0.9667*	*0.9554*	*0.865*	*0.9565*	*0.9665*
16	*250*	*83*	*28*	*185*	*22102.66*	*0.0431*	*0.9834*	*0.9631*	*0.778*	*0.9799*	*0.9677*

Table 2: Average Rank of Estimator Ranks based on the Mean Square Error

Information Case	**Rank of Mean Square Errors for the Estimator for the 16 populations.**																**Average Rank**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	
FIC	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	***1***
PIC-I	*4*	*4*	*4*	*4*	*3*	*4*	*4*	*3*	*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*	*2*	***3***
PIC-III	*2*	*2*	*2*	*2*	*2*	*2*	*1*	*1*	*2*	*2*	*2*	*2*	*2*	*2*	*2*	*3*	***2***
NIC	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	*5*	***5***

4.2: Conclusion

This research has used the method of [16] to propose two (2) partial information cases (PIC-II and PIC-III) in addendum to the existing three estimators established in [14]. The efficiency of the five estimators has been ascertained both theoretically and empirically. The estimator in FIC is efficient over all other estimators while estimator in NIC is less efficient to all other estimators. Estimator in PIC-III claims to be efficient over estimators in PIC-I and PIC-II. Similarly, estimator in PIC-II claims efficiency over estimator in PIC-I. It will be observed that the proposed estimators (PIC-II and PIC-III), in this study, are efficient over the PIC-I and NIC as suggested in [14].

Finally, this study has developed estimator schema for the easy understanding, modification and abridgement of lengthy estimators in two-phase sampling.

Reference

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

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

[3] Cochran, W.G. (1940). 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

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

[5] Tripathi T.P. (1970). Contributions of the Sampling Theory using multivariate information. Ph.D. thesis submitted to Punjabi University, Patiala, India.

[6] Das. A.K. (1988). 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.

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

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

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

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

[11] Rajesh S., Pankaj C., Nirmala S. and Florentins S. (2007). Ratio-Product type exponential estimator or estimating finite population mean using information on auxiliary attributes. Renaissance, High Press, USA.

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

[13] Moeen M., Shahbaz Q. and Hanif M. (2012). Mixture Ratio and Regression Estimators using multi-auxiliary variables and attributes in single phase sampling. World Applied Sciences Journal. Vol. 18, Pp. 1518-1526.

[14] Waweru, P.M., Kung'u, J and Kahiri, J. (2014). 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.

[15] Kung'u J., Chumba G. and Odongo L. (2014), 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.

[16] Ogunyinka P.I. and Sodipo A.A. (2017). 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. Vol. 17(1), ISSN: 0975-1262, pp. 1-20, http://scientificadvances.co.in

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

[18] Hamad Z. and Hanif M. (2010). Generalized Multivariate Ratio and Regression Estimators. LAP LAMBERT Academic Publishing, AG & Co. KG, ISBN:978-8383-4678-6.