

Generalized P-phased Regression Estimators with Single and Two Auxiliary Variables

Farhan Hameed

Punjab Higher Education Department

Research Scholar, Govt. College University (GCU), Lahore, Pakistan

qunchless@gmail.com

Hina Khan

Department of Statistics, Govt. College University (GCU), Lahore, Pakistan

hinakhan@gc.edu.pk

Abstract

Multiphase sampling has been the concept not being utilized is estimation of ratio and regression estimator widely. In the recent study we have proposed new dimension of sampling survey of estimations by proposing two generalized p-phase regression estimators with single and two auxiliary variables for estimating population mean. The proposed estimators are the generalized p-phase cases of Hanif et al (2015) and Hanif (2007) respectively. Both the estimators from which we took motivation are now special cases of our proposed estimators. We have derived unbiasedness, expression of Mean Square Errors along with family of estimators based upon p-phased generalization. We have derived expression of MSE in such a way that these expression can be used to obtained results for every phase we desire. By conducting empirical study on proposed estimators we have shown many situation in which MSE can be reduced by increasing number of phases. Hence, our study will open new horizon in the field of multiphase sampling where a lot of challenges are waiting to be resolved by proposing new estimators for phases above 2nd phase.

Keywords: MSE, NIC, Generalized P-Phased, Auxiliary Variable.

1. Introduction

Sampling survey is perhaps the oldest statistical procedure to determine the accurate and useful estimates under prevailing constraints of time and money. The regression and ratio methods of estimation are two strongest pillars of sampling survey. In ratio and regression estimation many interventions in terms of estimators with different structural and functional form have been made. Two phase and multiphase sampling are the concepts associated in estimation of population mean from finite population under different cases of availability or non- availability of auxiliary information. In the recent study we have proposed new dimension of sampling survey of estimations by proposing generalized p-phase regression estimator. The proposed estimators undertakes multiphase sampling producers for estimation. Our estimator is attempt to utilize information from phases above second phase. Not only we become able to gather maximum information about auxiliary variables, but also in many cases the Mean Square Error (MSE) is also reduced.

The regression estimator have been used widely by Srivastava (1967), Walsh (1970), Reddy (1973, 74), Gupta (1978), Sahai (1979), Vos (1980) for estimation. The Classical Regression Estimator (CRE) is

$$T_1 = \bar{y} + \beta_{yx}(\bar{X} - \bar{x}) \quad (1.1)$$

Vos (1980) used weighted average of Mean Per Unit (MPU) Ratio Estimator to propose following estimator:-

$$T_2 = \alpha \bar{y} + (1 - \alpha) \frac{\bar{y}}{\bar{x}} \bar{X} \quad (1.2)$$

where α is constant. The estimator (1.8.1.3) is identical to MPU estimator for $\alpha = 1$ and is identical to RE for $\alpha = 0$.

Das (1988) with little modification of CRE proposed following estimator:

$$T_3 = \bar{y} - W\lambda(\bar{x} - \bar{X}) \quad (1.3)$$

For $W = 1$ estimator (1.3) reduces to (1.2) and for $W = 0$ it reduces to MPUE.

Mohanty (1967) combined Ratio Estimator with two auxiliary variables and proposed the estimator:-

$$T_4 = \bar{y} + \beta_{yx}(\bar{X} - \bar{x}) \frac{\bar{Z}}{\bar{z}} \quad (1.4)$$

The estimator (1.4) reduces to (1.1) under non-availability of Z. The estimator (1.4) was termed as Regression-In-Ratio Estimator (RIRE).

Samiuddin and Hanif (2006) presented RE in single phase as:-

$$T_5 = \bar{y} + \alpha(\bar{X} - \bar{x}) + \beta(\bar{Z} - \bar{z}) \quad (1.5)$$

Now, some of the estimators in two phase sampling presented in literature are as follows:-

Mohanty (1967) followed (1.4) to proposed the estimator in 2nd phase sampling with two auxiliary variables under No Information Case (NIC):-

$$T_6 = [\bar{y}_2 + b_{yx}(\bar{x}_1 - \bar{x}_2)] \frac{\bar{Z}_1}{\bar{z}_1} \quad (1.6)$$

Chand (1975) presented estimator in Partial Information Case (PIC):-

$$T_7 = \frac{\bar{y}_2}{\bar{z}_2} [\bar{z}_1 + b_{zx}(\bar{X} - \bar{x}_1)] \quad (1.7)$$

There have been numerous estimators in the literature other. Some notable references are Khaire and Srivastava (1981), Saho and Saho (1993), Upadhyaya and Singh (2001), Roy (2003), Samiuddin and Hanif (2006), Hanif et al (2010).

The attempts to present RE above second phase is a new dimension in this context Hanif et al (2015) presented two estimators in three phase and four phase sampling.

$$\bar{y}_{H_1} = \bar{y}_3 + \beta_1(\bar{x}_1 - \bar{x}_2) + \beta_2(\bar{x}_2 - \bar{x}_3) \quad (1.8)$$

$$\bar{y}_{H_2} = \bar{y}_4 + \beta_1(\bar{x}_1 - \bar{x}_2) + \beta_2(\bar{x}_2 - \bar{x}_3) + \beta_3(\bar{x}_3 - \bar{x}_4) \quad (1.9)$$

The optimizing constants $\beta_1, \beta_2, \beta_3$ were identical as under:-

$$\beta_1 = \beta_2 = \beta_3 = \frac{S_{xy}}{S_x^2}$$

The corresponding MSE of (1.8) and (1.9) were

$$MSE(\bar{y}_{H_1}) = \theta_3 S_y^2 (1 - \rho^2) + \theta_1 \rho^2 S_y^2 \quad (1.10)$$

The corresponding expression of MSE of (3.7.2) is as under:-

$$MSE(\bar{y}_{H_2}) = \theta_4 S_y^2 (1 - \rho^2) + \theta_1 \rho^2 S_y^2 \quad (1.11)$$

Based upon the finding in third and fourth phase Hanif et al (2015) concluded that MSE will tends to increase. Earlier, Samiuddin and Hanif (2007) considered NIC to present RE in two phase sampling which was similar to structural formation of (1.8) and (1.9), but in two phase sampling with two auxiliary variables :-

$$\bar{y}_{SH} = \bar{y}_2 + \alpha(\bar{x}_1 - \bar{x}_2) + \beta(\bar{z}_1 - \bar{z}_2) \quad (1.12)$$

Here, the co-efficient α and β were obtained through optimization of the expression of MSE. The expression of MSE of (1.12) was obtained as:-

$$MSE(\bar{y}_{SH}) = \frac{\bar{Y}^2 C_y^2}{1 - \rho_{xz}^2} \left[\theta_2 (1 - \rho_{xy}^2 - \rho_{yz}^2 - \rho_{xz}^2 + 2\rho_{xy}\rho_{xz}\rho_{yz}) + \theta_1 (\rho_{xy}^2 + \rho_{yz}^2 - 2\rho_{xy}\rho_{xz}\rho_{yz}) \right] \quad (1.13)$$

The optimizing coefficients were:-

$$\alpha = \frac{\bar{Y} C_y (\rho_{xy} - \rho_{xz} \rho_{yz})}{\bar{X} C_y (1 - \rho_{xz}^2)} \quad (1.14)$$

$$\beta = \frac{\bar{Y} C_y (\rho_{yz} - \rho_{xz} \rho_{xy})}{\bar{Z} C_z (1 - \rho_{xz}^2)} \quad (1.15)$$

2. P-Phased Sampling Notations

The p-phase sampling plan is just the extension of 2nd phase sampling procedure. Under such plan subsequent sub samples of sizes n_1, n_2, \dots, n_p are drawn for first phase, 2nd phase up to pth phase respectively. The information on auxiliary and understudy variables will be gather in similar pattern as in two phase case.

Let Y_i, X_i and Z_i are variables with means \bar{Y}, \bar{X} and \bar{Z} and variances S_y^2, S_x^2 and S_z^2 respectively. $C_y = \frac{S_y}{\bar{Y}}$, $C_x = \frac{S_x}{\bar{X}}$ and $C_z = \frac{S_z}{\bar{Z}}$ are co-efficient of variations. Also ρ_{xy}, ρ_{xz} and ρ_{yz} will represent population correlation coefficients between X & Y , X & Z and Y & Z respectively. $n_i, i = 1, 2, \dots, p$, are subsequent sample sizes. Whereas, \bar{x}_i, \bar{y}_i and $\bar{z}_i, i = 1, 2, \dots, p$, represent sample means at first, second and so on phases.

The Finite Population Correction Factor (FPC) in p-phase Simple Random Sampling With Out Replacement (SRWOR) sampling will be denoted by $\theta_i, i = 1, 2, \dots, p$, and given as:-

$$\theta_i = \frac{N - n_i}{N n_i} = \frac{N}{N n_i} - \frac{n_i}{N n_i} = \frac{1}{n_i} - \frac{1}{N} = n_i^{-1} - N^{-1} \quad (2.1)$$

Also we have:-

$$\theta_i > \theta_j \text{ for } i > j \quad (2.2)$$

The error term in each of the variables will be represented as under with the assumption that the quantities $|\bar{e}_{x_i}|$, $|\bar{e}_{z_i}|$ are very small as compare to $|\bar{X}|$ and $|\bar{Z}|$.

$$\left. \begin{aligned} \bar{x}_i &= \bar{X} + \bar{e}_{x_i} \\ \bar{z}_i &= \bar{Z} + \bar{e}_{z_i} \\ \bar{y}_i &= \bar{Y} + \bar{e}_{y_i} \\ i &= 1, 2, \dots, p \end{aligned} \right\} \quad (2.3)$$

Furthermore, the following results for expectations can be derived using (2.3)

$$\left. \begin{aligned} E(\bar{e}_{x_i}) &= E(\bar{x}_i - \bar{X}) = E(\bar{x}_i) - E(\bar{X}) = \bar{X} - \bar{X} = 0 \\ E(\bar{e}_{z_i}) &= E(\bar{z}_i - \bar{Z}) = E(\bar{z}_i) - E(\bar{Z}) = \bar{Z} - \bar{Z} = 0 \\ E(\bar{e}_{y_i}) &= E(\bar{y}_i - \bar{Y}) = E(\bar{y}_i) - E(\bar{Y}) = \bar{Y} - \bar{Y} = 0 \\ i &= 1, 2, 3, \dots, p \end{aligned} \right\} \quad (2.4)$$

The expectation of square of error terms will be as under:-

$$\left. \begin{aligned} E(\bar{e}_{x_i})^2 &= E(\bar{x}_i - \bar{X})^2 = \theta_i S_x^2 = \theta_i \bar{X}^2 C_x^2 \\ E(\bar{e}_{z_i})^2 &= E(\bar{z}_i - \bar{Z})^2 = \theta_i S_z^2 = \theta_i \bar{Z}^2 C_z^2 \\ E(\bar{e}_{y_i})^2 &= E(\bar{y}_i - \bar{Y})^2 = \theta_i S_y^2 = \theta_i \bar{Y}^2 C_y^2 \\ i &= 1, 2, 3, \dots, p \end{aligned} \right\} \quad (2.5)$$

In addition to (2.4) and (2.5) the following results of expectations from cross products will also be utilized:-

$$\left. \begin{aligned} E(\bar{e}_{x_i} \cdot \bar{e}_{x_j}) &= \theta_i \bar{X}^2 C_x^2 \\ E(\bar{e}_{z_i} \cdot \bar{e}_{z_j}) &= \theta_i \bar{Z}^2 C_z^2 \\ E(\bar{e}_{y_i} \cdot \bar{e}_{y_j}) &= \theta_i \bar{Y}^2 C_y^2 \\ E(\bar{e}_{x_i} \cdot \bar{e}_{z_j}) &= \theta_i S_{xz} = \theta_i \bar{X} \bar{Z} \rho_{xz} C_x C_z \\ E(\bar{e}_{x_i} \cdot \bar{e}_{y_j}) &= \theta_i S_{xy} = \theta_i \bar{X} \bar{Y} \rho_{xy} C_x C_y \\ E(\bar{e}_{z_i} \cdot \bar{e}_{y_j}) &= \theta_i S_{yz} = \theta_i \bar{Z} \bar{Y} \rho_{zy} C_z C_y \\ i \leq j, \quad i &= 1, 2, 3, \dots, p \end{aligned} \right\} \quad (2.6)$$

Finally, the following results will also be used:-

$$\left. \begin{aligned} E(\bar{e}_{x_i} - \bar{e}_{x_j})^2 &= (\theta_j - \theta_i) \bar{X}^2 C_x^2 \\ E(\bar{e}_{z_i} - \bar{e}_{z_j})^2 &= (\theta_j - \theta_i) \bar{Z}^2 C_z^2 \\ E(\bar{e}_{x_i} - \bar{e}_{x_{i+1}}) \bar{e}_{y_p} &= (\theta_i - \theta_{i+1}) S_{xy} = (\theta_i - \theta_{i+1}) \bar{X} \bar{Y} \rho_{xy} C_x C_y \\ E(\bar{e}_{z_i} - \bar{e}_{z_{i+1}}) \bar{e}_{y_p} &= (\theta_i - \theta_{i+1}) S_{zy} = (\theta_i - \theta_{i+1}) \bar{Z} \bar{Y} \rho_{zy} C_z C_y \\ i < j, \quad i &= 1, 2, \dots, p \end{aligned} \right\} \quad (2.7)$$

3. Proposed Estimators

We took motivation from Hanif et al (2015) and Samiuddin and Hanif (2006) to present estimators in generalized p-phased estimators as under. Generalizing Hanif et al (2015) estimators (1.8) and (1.9) we propose.

(a)- Proposed Estimator-I

Taking Motivation from Hanif et al (2015), we propose the following generalized p-phased regression estimator with single auxiliary variable.

$$\bar{y}_{F_1} = \bar{y}_p + \left. \sum_{i=1}^{p-1} \beta_i (\bar{x}_i - \bar{x}_{i+1}) \right\} \quad p \geq 2 \quad (3.1)$$

(b) Proposed Estimator-II

Similarly, the Samiuddin and Hanif (2006) estimator presented in (1.12) is p-phased generalized as:-

$$\bar{y}_{F_2} = \bar{y}_p + \left. \sum_{i=1}^{p-1} \alpha_i (\bar{x}_i - \bar{x}_{i+1}) + \sum_{i=1}^{p-1} \beta_i (\bar{z}_i - \bar{z}_{i+1}) \right\} \quad p \geq 2 \quad (3.2)$$

4. Unbiasedness and MSE of proposed Estimator-I

Now, we will prove unbiasedness of the proposed estimators (3.1) and (3.2) as well as the expression of MSE will be derived.

Considering (3.1) and using comment (2.3) we get:

$$\bar{y}_{F_1} - \bar{Y} = \bar{e}_{y_p} + \sum_{i=1}^{p-1} \beta_i (\bar{e}_{x_i} - \bar{e}_{x_{i+1}}) \quad (4.1)$$

Applying expectation and using (2.4) we obtained:-

$$\begin{aligned} E(\bar{y}_{F_1}) - \bar{Y} &= 0 \\ E(\bar{y}_{F_1}) &= \bar{Y} \end{aligned} \quad (4.2)$$

From (4.2) we observed that (3.1) is unbiased estimator of Population Mean.

Now, to derive MSE of (3.1) consider (4.1) and squaring both sides and applying expectations

$$\begin{aligned} E[(\bar{y}_{F_1} - \bar{Y})^2] &= E\left[(\bar{e}_{y_p})^2\right] + E\left[\sum_{i=1}^{p-1} \beta_i^2 (\bar{e}_{x_i} - \bar{e}_{x_{i+1}})^2 + 2 \sum_{i < j=1}^{p-1} \beta_i \beta_j (\bar{e}_{x_i} - \bar{e}_{x_{i+1}})(\bar{e}_{x_j} - \bar{e}_{x_{j+1}})\right] \\ &\quad + 2 E\left[\sum_{i=1}^{p-1} \beta_i (\bar{e}_{x_i} - \bar{e}_{x_{i+1}}) \bar{e}_{y_p}\right] \end{aligned} \quad (4.3)$$

Upon using (2.6) the third term on R.H.S of (4.3) will vanish and with results (2.6) and (2.7) we obtain:-

$$MSE(\bar{y}_{F_1}) = \theta_p S_y^2 + \sum_{i=1}^{p-1} \beta_i^2 (\theta_{i+1} - \theta_i) S_x^2 + 2 \sum_{i=1}^{p-1} \beta_i (\theta_i - \theta_{i+1}) S_{xy} \quad (4.4)$$

The optimizing co-efficient β_i will be determine by minimization of (4.4). For this purpose we proceed as:

$$\begin{aligned} \frac{\partial MSE(\bar{y}_{F_1})}{\partial \beta_i} &= \frac{\partial}{\partial \beta_i} (\theta_p S_y^2) + \frac{\partial}{\partial \beta_i} \sum_{i=1}^{p-1} \beta_i^2 (\theta_{i+1} - \theta_i) S_x^2 + 2 \frac{\partial}{\partial \beta_i} \sum_{i=1}^{p-1} \beta_i (\theta_i - \theta_{i+1}) S_{xy} \\ \frac{\partial MSE(\bar{y}_{F_1})}{\partial \beta_i} &= 2 \sum_{i=1}^{p-1} \beta_i (\theta_{i+1} - \theta_i) S_x^2 + 2 \sum_{i=1}^{p-1} (\theta_i - \theta_{i+1}) S_{xy} \end{aligned} \quad (4.5)$$

Now putting:-

$$\frac{\partial MSE(\bar{y}_{F_1})}{\partial \beta_i} = 0$$

We get:-

$$2 \sum_{i=1}^{p-1} (\theta_{i+1} - \theta_i) [\beta_i S_x^2 - S_{xy}] = 0$$

Upon Simplification:

$$\beta_i = \frac{S_{xy}}{S_x^2} \quad (4.6)$$

From (4.6), $\beta_1, \beta_2, \beta_3 \dots \dots \dots \beta_{p-1}$ are identical i.e.

$$\beta_i = \frac{S_{xy}}{S_x^2} = \beta(\text{say}), i = 1, 2, 3, \dots \dots \dots (p-1) \quad (4.7)$$

Now, differentiate (4.2.1.5) again w.r.t β_i

$$\frac{\partial^2 MSE(\bar{y}_{F_1})}{\partial \beta_i^2} = 2 S_x^2 (\theta_p - \theta_1) \quad (4.8)$$

The expression (4.8) will turn out to be positive because that S_x^2 is variance and the term $(\theta_p - \theta_1)$ will also be positive for $= 2, 3, \dots \dots \dots p$, under the condition presented in (2.2). Furthermore, because, these are non-zero quantities so the second derivative test will not fall into indecision zone. Now, we replace the optimized values obtained in (4.7) into (4.4) to get the final expression of MSE of (3.1).

$$MSE(\bar{y}_{F_1}) = \theta_p S_y^2 + \sum_{i=1}^{p-1} \beta [\beta (\theta_{i+1} - \theta_i) S_x^2 + 2(\theta_i - \theta_{i+1}) S_{xy}] \quad (4.9)$$

Expressing (4.7) as

$$\beta = \frac{S_{xy}S_y}{S_x^2S_y} = \frac{S_y}{S_x} \cdot \frac{S_{xy}}{\sqrt{S_x^2S_y^2}} = \frac{S_y}{S_x} \cdot \rho_{xy} \quad (4.10)$$

and using in (4.9)

$$MSE(\bar{y}_{F_1}) = \theta_p S_y^2 + S_y^2 \cdot \rho_{xy}^2 \sum_{i=1}^{p-1} [\theta_i - \theta_{i+1}] \quad (4.10)$$

On expanding the summation we get:-

$$MSE(\bar{y}_{F_1}) = \theta_p S_y^2 (1 - \rho_{xy}^2) + \theta_1 S_y^2 \cdot \rho_{xy}^2 \quad (4.11)$$

Hence, (4.11) provides generalized p-phased expression of proposed estimator (3.1). From the expression (4.11) we observed that our proposed estimator is generalized case of Hanif et al (2015) estimators (1.8) and (1.9). Based upon (4.11) proposed estimator (3.1) can be considered as generalized p-phased regression estimator with single auxiliary variable having family of estimators for $p \geq 2$.

Table 1: Family of proposed Estimator-1

p	Estimator($\bar{y}_{F_{1(p)}}$)	MSE($\bar{y}_{F_{1(p)}}$)
≥ 2	$\bar{y}_{F_{1(p)}} = \bar{y}_p + \sum_{i=1}^{p-1} \beta_i (\bar{x}_i - \bar{x}_{i+1})$ <p style="text-align: right;">From (3.1)</p>	$\theta_p S_y^2 (1 - \rho_{xy}^2) + \theta_1 S_y^2 \cdot \rho_{xy}^2$ <p style="text-align: right;">From (4.11)</p>
3	$\bar{y}_{F_{1(3)}} = \bar{y}_{H_1} = \bar{y}_3 + \beta_1 (\bar{x}_1 - \bar{x}_2) + \beta_2 (\bar{x}_2 - \bar{x}_3)$ <p style="text-align: right;">From (1.8)</p>	$\theta_3 S_y^2 (1 - \rho_{xy}^2) + \theta_1 S_y^2 \cdot \rho_{xy}^2$ <p style="text-align: right;">From (1.10)</p>
4	$\bar{y}_{F_{1(4)}} = \bar{y}_{H_2}$ $= \bar{y}_4 + \beta_1 (\bar{x}_1 - \bar{x}_2) + \beta_2 (\bar{x}_2 - \bar{x}_3) + \beta_3 (\bar{x}_3 - \bar{x}_4)$ <p style="text-align: right;">From (1.9)</p>	$\theta_4 S_y^2 (1 - \rho_{xy}^2) + \theta_1 S_y^2 \cdot \rho_{xy}^2$ <p style="text-align: right;">From (1.11)</p>

4.1 Intra-Phase Comparison for Proposed Estimator-I

Considering (4.11) and replacing p by $p - 1$, we get:-

$$MSE(\bar{y}_{F_1})_{p-1} = \theta_{p-1} S_y^2 (1 - \rho_{xy}^2) + \theta_1 S_y^2 \cdot \rho_{xy}^2 \quad (4.1.1)$$

Now, comparing (4.11) and (4.1.1):-

$$\begin{aligned} MSE(\bar{y}_{F_1})_p &< MSE(\bar{y}_{F_1})_{p-1} \\ \theta_p S_y^2 (1 - \rho_{xy}^2) + \theta_1 S_y^2 \cdot \rho_{xy}^2 &< \theta_{p-1} S_y^2 (1 - \rho_{xy}^2) + \theta_1 S_y^2 \cdot \rho_{xy}^2 \\ \theta_p &< \theta_{p-1} \end{aligned} \quad (4.1.2)$$

Condition (4.1.2) suggest that for every next phase FPC should decrease which is contradictory to basic methodology of multiphase sampling as we have stated in (2.2). Therefore, MSE will lean towards increasing pattern.

4.2 Empirical Study of Proposed Estimator-I

Now, we undertake the following population to observe the behavior of MSE of proposed estimator-I

Table 4.2.1: Summary of the Population

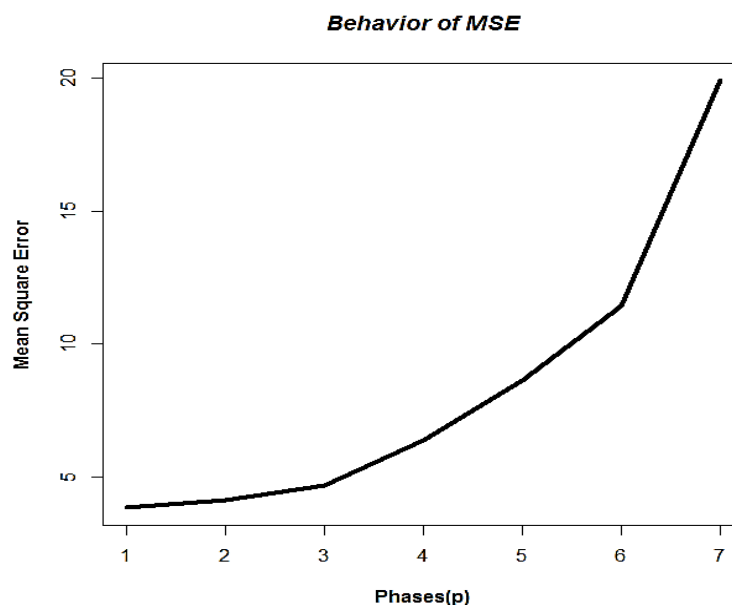
Understudy Variable	Auxiliary Variable	Population Size	Var(Y)	Var(X)	Correlation Co-efficient
Y= MPG of Car	X= Weight of Car	N=81	$S_Y^2 = 101.49$	$S_X^2=66.85$	$\rho_{XY} = -0.9128$

R-codes for the expression (4.11) are used to find MSE at different phases. Furthermore, intra phase relative efficiency is also computed. The finding are presented in the table 4.2.2 along with corresponding graphical depiction in Figure 4.2.1

Table 4.2.2: Behavior of MSE of Proposed Estimator-I

Sample Size	Phases	MSE	Relative Efficiency
20	1	3.82177	
15	2	4.10389	107.3819226
10	3	4.66813	113.7489112
5	4	6.36085	136.2612118
3	5	8.61781	135.4820582
2	6	11.439	132.7368607
1	7	19.9026	173.9889294

Figure 7.1.2: Behavior of MSE of Proposed Estimator-I

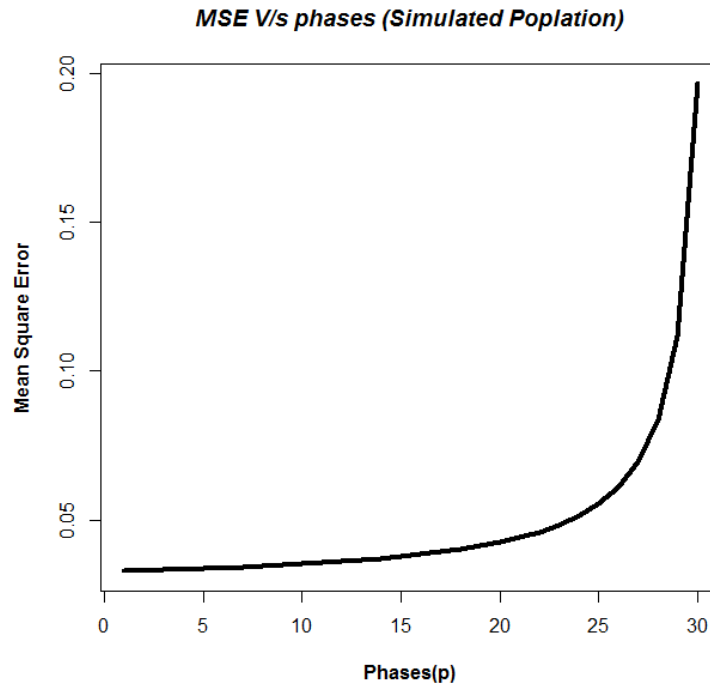


We can observe that the MSE tends to increase phase by phase. At earlier phases MSE shows slow and steady increase very little, but dynamic quick increasing pattern is observed afterwards. Relative efficient will always greater than 100% because of increasing pattern of MSE. Because of small population the curve in Figure 4.2.1 shows sharp slope. Now, simulating population using multivariate normal distribution up to $N = 200000$. First stage sample is taken as 3000, afterwards decrease of 100 is at each phase is considered until $p = 30$. The results are reflected in Figure 4.2.2. The curve is very smooth and steady in slope, but still upward.

We can conclude that if we have large population and considerably larger sample sizes, there is negligible danger of losing efficiency. As in the current example we can easily observe that up to phase 25 intra phase MSEs are adjacent. Therefore, proposed estimator (3.1) is fair choice in sufficiently large population.

Based upon the mathematical finding and empirical studies. Now, we are in a straight forward position to conclude that MSE of (3.1) will tend to increase by increasing phases. Such a statement is justified, because we have generalized p-phased analysis of the general case of Hanif et al (2015). They just utilized four phase to conclude that “MSE will increase should be the case”. Whereas, we not only conducted intra-phase mathematical comparison, but also supported our conclusion with the empirical study having flexible choice of phases

Figure 4.2.2: Behavior of MSE of Proposed Estimator with Simulation



5.1 Unbiasedness and MSE of Proposed Estimator-II

First we derive unbiasedness of (3.2), for this consider use (2.3) in (3.2) and applying expectations both sides:-

$$E(\bar{y}_{F_2} - \bar{Y}) = \left[E(\bar{e}_{y_p}) + \sum_{i=1}^{p-1} \alpha_i E(\bar{e}_{x_i} - \bar{e}_{x_{i+1}}) + \sum_{i=1}^{p-1} \beta_i E(\bar{e}_{z_i} - \bar{e}_{z_{i+1}}) \right] \quad (5.1.1)$$

From the results in (2.4) the R.H.S on (5.1.1) vanished and we get:-

$$E(\bar{y}_{F_2}) = \bar{Y} \quad (5.1.2)$$

(5.1.2) provides evidence for unbiasedness of proposed estimator-II.

Now, for MSE Estimator (3.2), we proceed as

$$\bar{y}_{F_2} - \bar{Y} = \bar{e}_{y_p} + \sum_{i=1}^{p-1} \alpha_i (\bar{e}_{x_i} - \bar{e}_{x_{i+1}}) + \sum_{i=1}^{p-1} \beta_i (\bar{e}_{z_i} - \bar{e}_{z_{i+1}})$$

Squaring both sides and applying expectations

$$\begin{aligned} MSE(\bar{y}_{F_2}) &= E(\bar{e}_{y_p})^2 + E \left[\sum_{i=1}^{p-1} \alpha_i^2 E(\bar{e}_{x_i} - \bar{e}_{x_{i+1}})^2 + \right. \\ &\quad \left. 2 \sum_{i < j=1}^{p-1} \alpha_i \alpha_j E(\bar{e}_{x_i} - \bar{e}_{x_{i+1}})(\bar{e}_{x_j} - \bar{e}_{x_{j+1}}) \right] \\ &\quad + E \left[\sum_{i=1}^{p-1} \beta_i^2 E(\bar{e}_{z_i} - \bar{e}_{z_{i+1}})^2 + \right. \\ &\quad \left. 2 \sum_{i < j=1}^{p-1} \beta_i \beta_j E(\bar{e}_{z_i} - \bar{e}_{z_{i+1}})(\bar{e}_{z_j} - \bar{e}_{z_{j+1}}) \right] \\ &\quad + 2 \sum_{i=1}^{p-1} \alpha_i E(\bar{e}_{x_i} - \bar{e}_{x_{i+1}})(\bar{e}_{y_p}) + 2 \sum_{i=1}^{p-1} \beta_i E(\bar{e}_{z_i} - \bar{e}_{z_{i+1}})(\bar{e}_{y_p}) \\ &\quad + 2 E \left\{ \sum_{i=1}^{p-1} \alpha_i (\bar{e}_{x_i} - \bar{e}_{x_{i+1}}) \right\} \left\{ \sum_{i=1}^{p-1} \beta_i (\bar{e}_{z_i} - \bar{e}_{z_{i+1}}) \right\} \end{aligned} \quad (5.1.3)$$

The last term on R.H.S of (5.1.3) can be expanded as:-

$$\begin{aligned} E \left[\sum_{i=1}^{p-1} \alpha_i (\bar{e}_{x_i} - \bar{e}_{x_{i+1}}) \sum_{i=1}^{p-1} \beta_i (\bar{e}_{z_i} - \bar{e}_{z_{i+1}}) \right] \\ = \sum_{j=1}^{p-1} \left[\sum_{i=1}^{p-1} \alpha_i E(\bar{e}_{x_i} - \bar{e}_{x_{i+1}}) \beta_j (\bar{e}_{z_j} - \bar{e}_{z_{j+1}}) \right] \end{aligned} \quad (5.1.4)$$

For $i = j$, using results (2.5) and (2.6):-

$$E \left[\sum_{i=1}^{p-1} \alpha_i (\bar{e}_{x_i} - \bar{e}_{x_{i+1}}) \sum_{i=1}^{p-1} \beta_i (\bar{e}_{z_i} - \bar{e}_{z_{i+1}}) \right] \\ = \sum_{i=1}^{p-1} \alpha_i \beta_i (\theta_{i+1} - \theta_i) \bar{X} \bar{Z} \rho_{xz} C_x C_z \quad (5.1.5)$$

For $i \neq j$ we have two possibilities for $i < j$, using results (2.5) and (2.6):-

$$E(\bar{e}_{x_i} - \bar{e}_{x_{i+1}})(\bar{e}_{z_j} - \bar{e}_{z_{j+1}}) = \theta_i S_{xz} - \theta_i S_{xz} - \theta_{i+1} S_{xz} + \theta_{i+1} S_{xz} \\ E(\bar{e}_{x_i} - \bar{e}_{x_{i+1}})(\bar{e}_{z_j} - \bar{e}_{z_{j+1}}) = 0 \quad (5.1.6)$$

Similarly, for $i > j$.

$$E(\bar{e}_{x_i} - \bar{e}_{x_{i+1}})(\bar{e}_{z_j} - \bar{e}_{z_{j+1}}) = 0 \quad (5.1.7)$$

Therefore, with (5.1.5), (5.1.6) and (5.1.7), (5.1.4) will become:-

$$E \left[\sum_{i=1}^{p-1} \alpha_i (\bar{e}_{x_i} - \bar{e}_{x_{i+1}}) \sum_{i=1}^{p-1} \beta_i (\bar{e}_{z_i} - \bar{e}_{z_{i+1}}) \right] = \sum_{i=1}^{p-1} \alpha_i \beta_i (\theta_{i+1} - \theta_i) \bar{X} \bar{Z} \rho_{xz} C_x C_z \quad (5.1.8)$$

Similar, by using (2.5), (2.6) and (2.7):-

$$E \left[2 \sum_{i < j=1}^{p-1} \alpha_i \alpha_j E(\bar{e}_{x_i} - \bar{e}_{x_{i+1}})(\bar{e}_{x_j} - \bar{e}_{x_{j+1}}) \right] = 0 \quad (5.1.9)$$

$$E \left[2 \sum_{i < j=1}^{p-1} \beta_i \beta_j (\bar{e}_{z_i} - \bar{e}_{z_{i+1}})(\bar{e}_{z_j} - \bar{e}_{z_{j+1}}) \right] = 0 \quad (5.1.10)$$

Therefore, with comments (2.4), (2.5), (2.6) and (2.7) and results obtained in (5.1.8), (5.1.9) and (5.1.10), (5.1.3) will:-

$$MSE(\bar{y}_{F_2}) = \theta_p \bar{Y}^2 C_y^2 + \sum_{i=1}^{p-1} \alpha_i^2 (\theta_{i+1} - \theta_i) \bar{X}^2 C_x^2 + \sum_{i=1}^{p-1} \beta_i^2 (\theta_{i+1} - \theta_i) \bar{Z}^2 C_z^2 \\ + 2 \sum_{i=1}^{p-1} \alpha_i (\theta_i - \theta_{i+1}) \bar{X} \bar{Y} \rho_{xy} C_x C_y + 2 \sum_{i=1}^{p-1} \beta_i (\theta_i - \theta_{i+1}) \bar{Z} \bar{Y} \rho_{zy} C_z C_y \\ + 2 \sum_{i=1}^{p-1} \alpha_i \beta_i (\theta_{i+1} - \theta_i) \bar{X} \bar{Z} \rho_{xz} C_x C_z \quad (5.1.11)$$

To determine α_i and β_i differentiate (5.1.11) with respect to α_i and β_i respectively.

$$\begin{aligned} \frac{\partial}{\partial \alpha_i} [MSE(\bar{y}_{F_2})] &= 2 \sum_{i=1}^{p-1} \alpha_i (\theta_{i+1} - \theta_i) \bar{X}^2 C_x^2 + 2 \sum_{i=1}^{p-1} (\theta_i - \theta_{i+1}) \bar{X} \bar{Y} \rho_{xy} C_x C_y \\ &\quad + 2 \sum_{i=1}^{p-1} \beta_i (\theta_{i+1} - \theta_i) \bar{X} \bar{Z} \rho_{xz} C_x C_z \end{aligned} \quad (5.1.12)$$

$$\begin{aligned} \frac{\partial}{\partial \beta_i} [MSE(\bar{y}_{F_2})] &= 2 \sum_{i=1}^{p-1} \beta_i (\theta_{i+1} - \theta_i) \bar{Z}^2 C_z^2 + 2 \sum_{i=1}^{p-1} (\theta_i - \theta_{i+1}) \bar{Z} \bar{Y} \rho_{zy} C_z C_y \\ &\quad + 2 \sum_{i=1}^{p-1} \alpha_i (\theta_{i+1} - \theta_i) \bar{X} \bar{Z} \rho_{xz} C_x C_z \end{aligned} \quad (5.1.13)$$

Equating (5.1.12) and (5.1.13) to zero

$$\sum_{i=1}^{p-1} (\theta_{i+1} - \theta_i) \bar{X} C_x [\bar{X} C_x \alpha_i + \bar{Z} \rho_{xz} C_z \beta_i] = \sum_{i=1}^{p-1} (\theta_i - \theta_{i+1}) \bar{X} \bar{Y} \rho_{xy} C_x C_y \quad (5.1.13)$$

$$\begin{aligned} &\sum_{i=1}^{p-1} (\theta_{i+1} - \theta_i) \bar{Z} C_z [\bar{Z} C_z \beta_i + \bar{X} \rho_{xz} C_x \alpha_i] \\ &= \sum_{i=1}^{p-1} (\theta_i - \theta_{i+1}) \bar{Z} \bar{Y} \rho_{zy} C_z C_y \end{aligned} \quad (5.1.14)$$

From (5.1.13) and (5.1.14) we get:-

$$A\alpha_i + B\beta_i = C \quad (5.1.15)$$

$$E\alpha_i + D\beta_i = F \quad (5.1.16)$$

Where:-

$$\left. \begin{aligned} A &= \bar{X} C_x, B = \bar{Z} \rho_{xz} C_z, C = \bar{Y} \rho_{xy} C_y \\ D &= \bar{Z} C_z, E = \bar{X} \rho_{xz} C_x, F = \bar{Y} \rho_{zy} C_y \end{aligned} \right\} \quad (5.1.17)$$

Solving (5.1.15) and (5.1.16) simultaneously:-

$$\alpha_i = \frac{CD - BF}{AD - BE} \quad (5.1.18), \quad \beta_i = \frac{CE - AF}{BE - AD} \quad (5.1.19)$$

Making replacements in (5.1.18) and (5.1.19) from (5.1.17):-

$$\alpha_i = \frac{\bar{Y} C_y [\rho_{xy} - \rho_{xz} \rho_{zy}]}{\bar{X} C_x [1 - \rho_{xz}^2]} \quad (5.1.20), \quad \beta_i = \frac{\bar{Y} C_y [\rho_{zy} - \rho_{xy} \rho_{xz}]}{\bar{Z} C_z [1 - \rho_{xz}^2]} \quad (5.1.21)$$

Differentiate (5.1.13) and (5.1.14) partially w.r.t α_i and β_i

$$\frac{\partial^2}{\partial \alpha_i^2} [MSE(\bar{y}_{F_2})] = \bar{X}^2 C_x^2 (\theta_p - \theta_1) > 0 \quad (5.1.22)$$

$$\frac{\partial^2}{\partial \beta_i^2} [MSE(\bar{y}_{F_2})] = \bar{Z}^2 C_z^2 (\theta_p - \theta_1) > 0 \quad (5.1.23)$$

(5.1.22) and (5.1.23) will be positive quantities because they are product of two non-negative and non-zero terms. That is product of variance and $(\theta_p - \theta_1) > 0$ will be positive. Therefore, the values obtained in (5.1.20) and (5.1.21) will minimize MSE.

From (5.1.20) and (5.1.21) we observed that

$$\alpha_1 = \alpha_2 = \alpha_3 = \dots \dots \alpha_{p-1} = \frac{\bar{Y} C_y [\rho_{xy} - \rho_{xz} \rho_{zy}]}{\bar{X} C_x [1 - \rho_{xz}^2]} = \alpha(\text{say}) \quad (5.1.24)$$

$$\beta_1 = \beta_2 = \beta_3 = \dots \dots \beta_{p-1} = \frac{\bar{Y} C_y [\rho_{zy} - \rho_{xy} \rho_{xz}]}{\bar{Z} C_z [1 - \rho_{xz}^2]} = \beta(\text{say}) \quad (5.1.25)$$

Now, replacing (5.1.24) and (5.1.25) in (5.1.11)

$$\begin{aligned} MSE(\bar{y}_{F_2}) &= \theta_p \bar{Y}^2 C_y^2 + \sum_{i=1}^{p-1} \alpha^2 (\theta_{i+1} - \theta_i) \bar{X}^2 C_x^2 + \sum_{i=1}^{p-1} \beta^2 (\theta_{i+1} - \theta_i) \bar{Z}^2 C_z^2 \\ &+ 2 \sum_{i=1}^{p-1} \alpha (\theta_i - \theta_{i+1}) \bar{X} \bar{Y} \rho_{xy} C_x C_y + 2 \sum_{i=1}^{p-1} \beta (\theta_i - \theta_{i+1}) \bar{Z} \bar{Y} \rho_{zy} C_z C_y \\ &+ 2 \sum_{i=1}^{p-1} \alpha \beta (\theta_{i+1} - \theta_i) \bar{X} \bar{Z} \rho_{xz} C_x C_z \\ MSE(\bar{y}_{F_2}) &= \theta_p \left[\bar{Y}^2 C_y^2 + \alpha^2 \bar{X}^2 C_x^2 + \beta^2 \bar{Z}^2 C_z^2 - 2 \alpha \bar{X} \bar{Y} \rho_{xy} C_x C_y - 2 \beta \bar{Z} \bar{Y} \rho_{zy} C_z C_y \right. \\ &\quad \left. + 2 \alpha \beta \bar{X} \bar{Z} \rho_{xz} C_x C_z \right] \\ &+ \theta_1 \left[2 \alpha \bar{X} \bar{Y} \rho_{xy} C_x C_y + 2 \beta \bar{Z} \bar{Y} \rho_{zy} C_z C_y - \alpha^2 \bar{X}^2 C_x^2 - \beta^2 \bar{Z}^2 C_z^2 \right. \\ &\quad \left. - 2 \alpha \beta \bar{X} \bar{Z} \rho_{xz} C_x C_z \right] \quad (5.1.26) \end{aligned}$$

Replacing values from (5.1.24) and (5.1.25)

$$MSE(\bar{y}_{F_2}) = \theta_p \left[\begin{aligned} &\bar{Y}^2 C_y^2 + \left(\frac{\bar{Y} C_y [\rho_{xy} - \rho_{xz} \rho_{zy}]}{\bar{X} C_x [1 - \rho_{xz}^2]} \right)^2 \bar{X}^2 C_x^2 \\ &\left(\frac{\bar{Y} C_y [\rho_{zy} - \rho_{xy} \rho_{xz}]}{\bar{Z} C_z [1 - \rho_{xz}^2]} \right)^2 \bar{Z}^2 C_z^2 \\ &- 2 \left(\frac{\bar{Y} C_y [\rho_{xy} - \rho_{xz} \rho_{zy}]}{\bar{X} C_x [1 - \rho_{xz}^2]} \right) \bar{X} \bar{Y} \rho_{xy} C_x C_y \\ &- 2 \left(\frac{\bar{Y} C_y [\rho_{zy} - \rho_{xy} \rho_{xz}]}{\bar{Z} C_z [1 - \rho_{xz}^2]} \right) \bar{Z} \bar{Y} \rho_{zy} C_z C_y \\ &+ 2 \left(\frac{\bar{Y} C_y [\rho_{xy} - \rho_{xz} \rho_{zy}]}{\bar{X} C_x [1 - \rho_{xz}^2]} \right) \left(\frac{\bar{Y} C_y [\rho_{zy} - \rho_{xy} \rho_{xz}]}{\bar{Z} C_z [1 - \rho_{xz}^2]} \right) \bar{X} \bar{Z} \rho_{xz} C_x C_z \end{aligned} \right]$$

$$\begin{aligned}
 & + \theta_1 \left[2 \left(\frac{\bar{Y} C_y [\rho_{xy} - \rho_{xz} \rho_{zy}]}{\bar{X} C_x [1 - \rho_{xz}^2]} \right) \bar{X} \bar{Y} \rho_{xy} C_x C_y + 2 \left(\frac{\bar{Y} C_y [\rho_{zy} - \rho_{xy} \rho_{xz}]}{\bar{Z} C_z [1 - \rho_{xz}^2]} \right) \bar{Z} \bar{Y} \rho_{zy} C_z C_y \right. \\
 & \quad - \left(\frac{\bar{Y} C_y [\rho_{xy} - \rho_{xz} \rho_{zy}]}{\bar{X} C_x [1 - \rho_{xz}^2]} \right)^2 \bar{X}^2 C_x^2 - \left(\frac{\bar{Y} C_y [\rho_{zy} - \rho_{xy} \rho_{xz}]}{\bar{Z} C_z [1 - \rho_{xz}^2]} \right)^2 \bar{Z}^2 C_z^2 \\
 & \quad \left. - 2 \left(\frac{\bar{Y} C_y [\rho_{xy} - \rho_{xz} \rho_{zy}]}{\bar{X} C_x [1 - \rho_{xz}^2]} \right) \left(\frac{\bar{Y} C_y [\rho_{zy} - \rho_{xy} \rho_{xz}]}{\bar{Z} C_z [1 - \rho_{xz}^2]} \right) \bar{X} \bar{Z} \rho_{xz} C_x C_z \right] \\
 \\
 & MSE(\bar{y}_{F_2}) = \frac{\bar{Y}^2 C_y^2 \theta_p}{(1 - \rho_{xz}^2)^2} \left[\begin{aligned} & (1 - \rho_{xz}^2)^2 - \rho_{xy}^2 + 2\rho_{xz} \rho_{zy} \rho_{xy} - \rho_{zy}^2 \\ & - 4\rho_{zx}^3 \rho_{zy} \rho_{xy} + 2\rho_{zx}^2 \rho_{xy}^2 + 2\rho_{zx}^2 \rho_{zy}^2 \\ & - \rho_{xy}^2 \rho_{xz}^2 - \rho_{xz}^2 \rho_{zy}^2 + 2\rho_{xz}^3 \rho_{zy} \rho_{xy} \end{aligned} \right] \\
 & \quad + \frac{\bar{Y}^2 C_y^2 \theta_1}{(1 - \rho_{xz}^2)^2} \left[\begin{aligned} & \rho_{xy}^2 + \rho_{zy}^2 - 2\rho_{xz} \rho_{zy} \rho_{xy} - 2\rho_{xy}^2 \rho_{xz}^2 - 2\rho_{zy}^2 \rho_{xz}^2 + 4\rho_{xz}^3 \rho_{zy} \rho_{xy} \\ & + \rho_{xy}^2 \rho_{xz}^2 + \rho_{xz}^2 \rho_{zy}^2 - 2\rho_{xz}^3 \rho_{zy} \rho_{xy} \end{aligned} \right]
 \end{aligned}$$

On further simplification we get:-

$$MSE(\bar{y}_{F_2}) = \frac{\bar{Y}^2 C_y^2}{(1 - \rho_{xz}^2)} \left[\begin{aligned} & \theta_p (1 - \rho_{xy}^2 - \rho_{zy}^2 - \rho_{xz}^2 + 2\rho_{xz} \rho_{zy} \rho_{xy}) \\ & + \theta_1 (\rho_{xy}^2 + \rho_{zy}^2 - 2\rho_{xz} \rho_{zy} \rho_{xy}) \end{aligned} \right] \quad (5.1.27)$$

Hence, (5.1.27) provides the generalized p-phased expression of MSE of proposed Estimator-II. The expression is generalized p-phased for p=2,3,4-----.

For proposed estimator (3.2) the family is constructed in Table 5.1.1 for different phases with single and double auxiliary variables. From the expression in (5.1.27), we can approach to expression (4.11), if we consider only single auxiliary variable x and β_i as zero in (3.2). Therefore, the proposed estimator-I in (3.1) is member of the family of proposed estimator-II. Furthermore, the estimators of Hanif et al (2015) and Samiuddin and Hanif (2007) are cases of proposed estimator-II (3.2).

Table 5.1.1: Family of Proposed Estimator-II

Condition	Estimator / MSE
$p \geq 2, \alpha_i \neq 0, \beta_i \neq 0$	$\bar{y}_{F_{2(p)}} = \bar{y}_p + \sum_{i=1}^{p-1} \alpha_i (\bar{x}_i - \bar{x}_{i+1}) + \sum_{i=1}^{p-1} \beta_i (\bar{z}_i - \bar{z}_{i+1})$ <p style="text-align: right;">From (3.2)</p>
$MSE(\bar{y}_{F_{2(p)}}) =$	$\frac{\bar{Y}^2 C_y^2}{(1 - \rho_{xz}^2)} \left[\theta_p (1 - \rho_{xy}^2 - \rho_{zy}^2 - \rho_{xz}^2 + 2\rho_{xz} \rho_{zy} \rho_{xy}) + \theta_1 (\rho_{xy}^2 + \rho_{zy}^2 - 2\rho_{xz} \rho_{zy} \rho_{xy}) \right]$ <p style="text-align: right;">From (5.1.27)</p>
$p = 2, \alpha_i \neq 0, \beta_i \neq 0$	$\bar{y}_{F_{2(2)}} = \bar{y}_{SH} = \bar{y}_2 + \alpha (\bar{x}_1 - \bar{x}_2) + \beta (\bar{z}_1 - \bar{z}_2)$ <p style="text-align: right;">From (1.12)</p>
$MSE(\bar{y}_{F_{2(2)}} = \bar{y}_{SH}) =$	$\frac{\bar{Y}^2 C_y^2}{(1 - \rho_{xz}^2)} \left[\theta_2 (1 - \rho_{xy}^2 - \rho_{zy}^2 - \rho_{xz}^2 + 2\rho_{xz} \rho_{zy} \rho_{xy}) + \theta_1 (\rho_{xy}^2 + \rho_{zy}^2 - 2\rho_{xz} \rho_{zy} \rho_{xy}) \right]$
$p = 3, \alpha_i \neq 0, \beta_i \neq 0$	$\bar{y}_{F_{2(3)}} = \bar{y}_3 + \sum_{i=1}^2 \alpha_i (\bar{x}_i - \bar{x}_{i+1}) + \sum_{i=1}^2 \beta_i (\bar{z}_i - \bar{z}_{i+1})$
$MSE(\bar{y}_{F_{6(3)}}) =$	$\frac{\bar{Y}^2 C_y^2}{(1 - \rho_{xz}^2)} \left[\theta_3 (1 - \rho_{xy}^2 - \rho_{zy}^2 - \rho_{xz}^2 + 2\rho_{xz} \rho_{zy} \rho_{xy}) + \theta_1 (\rho_{xy}^2 + \rho_{zy}^2 - 2\rho_{xz} \rho_{zy} \rho_{xy}) \right]$
$p = 4, \alpha_i \neq 0, \beta_i \neq 0$	$\bar{y}_{F_{2(4)}} = \bar{y}_4 + \sum_{i=1}^3 \alpha_i (\bar{x}_i - \bar{x}_{i+1}) + \sum_{i=1}^3 \beta_i (\bar{z}_i - \bar{z}_{i+1})$
$MSE(\bar{y}_{F_{2(4)}}) =$	$\frac{\bar{Y}^2 C_y^2}{(1 - \rho_{xz}^2)} \left[\theta_4 (1 - \rho_{xy}^2 - \rho_{zy}^2 - \rho_{xz}^2 + 2\rho_{xz} \rho_{zy} \rho_{xy}) + \theta_1 (\rho_{xy}^2 + \rho_{zy}^2 - 2\rho_{xz} \rho_{zy} \rho_{xy}) \right]$
$p \geq 2, \alpha_i \neq 0, \beta_i = 0$	$\bar{y}_{F_{2(p)}} = \bar{y}_{F_{1(p)}} = \bar{y}_p + \sum_{i=1}^{p-1} \alpha_i (\bar{x}_i - \bar{x}_{i+1})$ <p style="text-align: right;">From (3.1)</p>
$MSE(\bar{y}_{F_{2(p)}} = \bar{y}_{F_{1(p)}}) =$	$\theta_p S_y^2 (1 - \rho_{xy}^2) + \theta_1 S_y^2 \cdot \rho_{xy}^2$ <p style="text-align: right;">From (4.1.11)</p>
$p = 3, \alpha_i \neq 0, \beta_i = 0$	$\bar{y}_{F_{6(3)}} = \bar{y}_{F_{1(3)}} = \bar{y}_{H_1} = \bar{y}_3 + \alpha_1 (\bar{x}_1 - \bar{x}_2) + \alpha_2 (\bar{x}_2 - \bar{x}_3)$ <p style="text-align: right;">From (1.8)</p>
$MSE(\bar{y}_{F_{2(3)}} = \bar{y}_{F_{1(3)}} = \bar{y}_{H_1}) =$	$\theta_3 S_y^2 (1 - \rho_{xy}^2) + \theta_1 S_y^2 \cdot \rho_{xy}^2$ <p style="text-align: right;">From (1.10)</p>
$p = 4, \alpha_i \neq 0, \beta_i = 0$	$\bar{y}_{F_{6(4)}} = \bar{y}_{F_{1(4)}} = \bar{y}_{H_2} = \bar{y}_4 + \sum_{i=1}^3 \alpha_i (\bar{x}_i - \bar{x}_{i+1})$ <p style="text-align: right;">From (1.9)</p>
$MSE(\bar{y}_{F_{2(4)}} = \bar{y}_{F_{1(4)}} = \bar{y}_{H_2}) =$	$\theta_4 S_y^2 (1 - \rho_{xy}^2) + \theta_1 S_y^2 \cdot \rho_{xy}^2$ <p style="text-align: right;">From (1.11)</p>

5.2 Empirical Study for Proposed Estimator-II

To perform empirical study on proposed estimator –II we will use different combination of correlations between the variable to demonstrate the pattern of MSE. Since we have two auxiliaries and an understudy variable therefore, we will have eight different combinations of correlations with respect to their signs. Such combinations are presented in Table 5.2.1 for moderate and low correlations.

Table 5.2.1: Combinations of Low-Moderate Level of Correlation

S.N	Combinations	\bar{Y}	ρ_{xy}	ρ_{xz}	ρ_{yz}	N
1	$\rho_{xy} < 0, \rho_{xz} > 0, \rho_{yz} < 0$	48.0557	-0.4406	0.80354	-0.3547	1678
2	$\rho_{xy} > 0, \rho_{xz} > 0, \rho_{yz} > 0$	48.0557	0.4406	0.80354	0.3547	1678
3	$\rho_{xy} < 0, \rho_{xz} < 0, \rho_{yz} < 0$	48.0557	-0.4406	-0.80354	-0.3547	1678
4	$\rho_{xy} > 0, \rho_{xz} > 0, \rho_{yz} < 0$	48.0557	0.4406	0.80354	-0.3547	1678
5	$\rho_{xy} < 0, \rho_{xz} < 0, \rho_{yz} > 0$	48.0557	-0.4406	-0.80354	0.3547	1678
6	$\rho_{xy} > 0, \rho_{xz} < 0, \rho_{yz} < 0$	48.0557	0.4406	-0.80354	-0.3547	1678
7	$\rho_{xy} < 0, \rho_{xz} > 0, \rho_{yz} > 0$	48.0557	-0.4406	0.80354	0.3547	1678
8	$\rho_{xy} > 0, \rho_{xz} < 0, \rho_{yz} > 0$	48.0557	0.4406	-0.80354	0.3547	1678

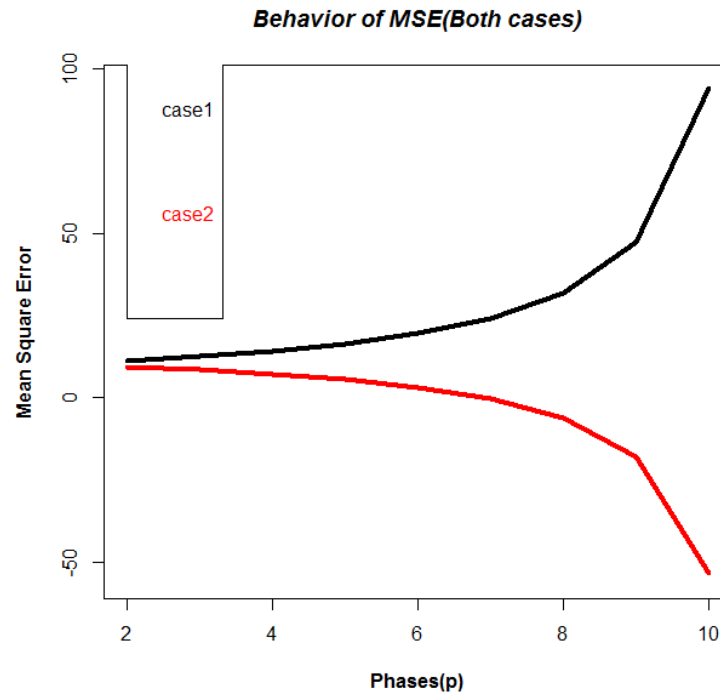
With combinations in Table 5.2.1 the MSE results are reported in Table 5.2.2. By observing results of MSE we can see only two type of results. Therefore, just two cases:- (i)-All correlations are positive or two out of three correlations are negative. (ii)-All correlations are negative or two out of three are positive are produced.

Now, conversing over the results in first case, MSE tends to increase slowly till phase 7. Afterwards, it has a sharp upward slop. For the seconds case MSE decreases steadily up to phase 6 and at phase 7 it is almost zero. We also observed that MSE produced are negative. This phenomenon is not the contradiction because, we have used population parameters to find values of MSE. The population parameters are constants whereas, FPC is variable depend upon phases. So, there is likelihood that MSE can be negative. Therefore, producing negative MSE is not objectionable.

The graphical pattern of both cases is under Figure 5.2.1. It is evident that for case (ii) MSE is reduced while phases increased.

Table 5.2.2: $MSE(\bar{y}_{F_{2(p)}})$ at Combinations of Low-Moderate Correlation

Sample Size	Phase	$MSE(\bar{y}_{F_{2(p)}})$							
		S.N#1	S.N#2	S.N#3	S.N#4	S.N#5	S.N#6	S.N#7	S.N#8
200	2	11.20	11.20	9.38	9.38	11.20	11.20	9.38	9.38
180	3	12.49	12.49	8.40	8.40	12.49	12.49	8.40	8.40
160	4	14.15	14.15	7.14	7.14	14.15	14.15	7.14	7.14
140	5	16.37	16.37	5.46	5.46	16.37	16.37	5.46	5.46
120	6	19.47	19.47	3.10	3.10	19.47	19.47	3.10	3.10
100	7	24.12	24.12	-0.42	-0.42	24.12	24.12	-0.42	-0.42
80	8	31.88	31.88	-6.31	-6.31	31.88	31.88	-6.31	-6.31
60	9	47.38	47.38	-18.09	-18.09	47.38	47.38	-18.09	-18.09
40	10	93.91	93.91	-53.41	-53.41	93.91	93.91	-53.41	-53.41

Figure 5.2.1: $MSE(\bar{y}_{F_{2(p)}})$ at Combinations of Low-Moderate Correlation

As we have just shown that only two types of results can be produced for all eight combinations of correlations therefore, in Table 5.2.3 we present intra-phase and second phased reference relative efficiencies

Table 5.2.3: Intra-Phase Relative Efficiency of $\bar{y}_{F_{2(p)}}$ at Moderate-Level Correlations

Phase	MSE(Case1)	MSE(Case2)	Each phase/2nd phase (%)		Each Phase / Previous Phase (%)	
			Case1	Case2	Case1	Case2
2	11.2	9.38	-	-	-	-
3	12.49	8.4	111.52	89.55	111.52	89.55
4	14.15	7.14	126.34	76.12	113.29	85.00
5	16.37	5.46	146.16	58.21	115.69	76.47
6	19.47	3.1	173.84	33.05	118.94	56.78
7	24.12	-0.42	215.36	-4.48	123.88	-13.55
8	31.88	-6.31	284.64	-67.27	132.17	1502.38
9	47.38	-18.09	423.04	-192.86	148.62	286.69
10	93.91	-53.41	838.48	-569.40	198.21	295.25

For case one when see that all the reported figures are above 100 which means that there is increasing pattern in MSE. The efficiency of the estimator reduces as phase increase. But there is not a much of difference as we move from one phase to another. For case 2 both types of efficiencies are less than 100 up to phase four. Having relative efficiency less than 100 means that the performance of the estimator is getting better and better with the increment in the phase. For example if we consider phase four in 2nd case the relative efficiency for current phase versus 2nd phase is 33.05 which means MSE at phase four is just 33% of MSE which was produced by phase two. Similarly, the intra-phase efficiency for case two is also better as with increasing phase MSE decreases resulting in better performance.

Now, we consider high correlation between variables and will observe what pattern MSE will display. For this purpose we will now consider only two combinations i.e. all positives and all negatives. Consider the following table 5.2.4

Table 5.2.4: Combinations of High Levels of Correlation

S.N	Combinations of Correlation	\bar{Y}	ρ_{xy}	ρ_{xz}	ρ_{yz}	N
1	High positive between understudy and auxiliaries Low positive between auxiliaries	48.0557	0.9406	0.20354	0.9547	1678
2	High positive between understudy and auxiliaries high positive between auxiliaries	48.0557	0.9406	0.90354	0.9547	1678
3	High negative between understudy and auxiliaries Low negative between auxiliaries	48.0557	-0.9406	-0.20354	-0.9547	1678
4	High negative between understudy and auxiliaries High negative between auxiliaries	48.0557	-0.9406	-0.90354	-0.9547	1678

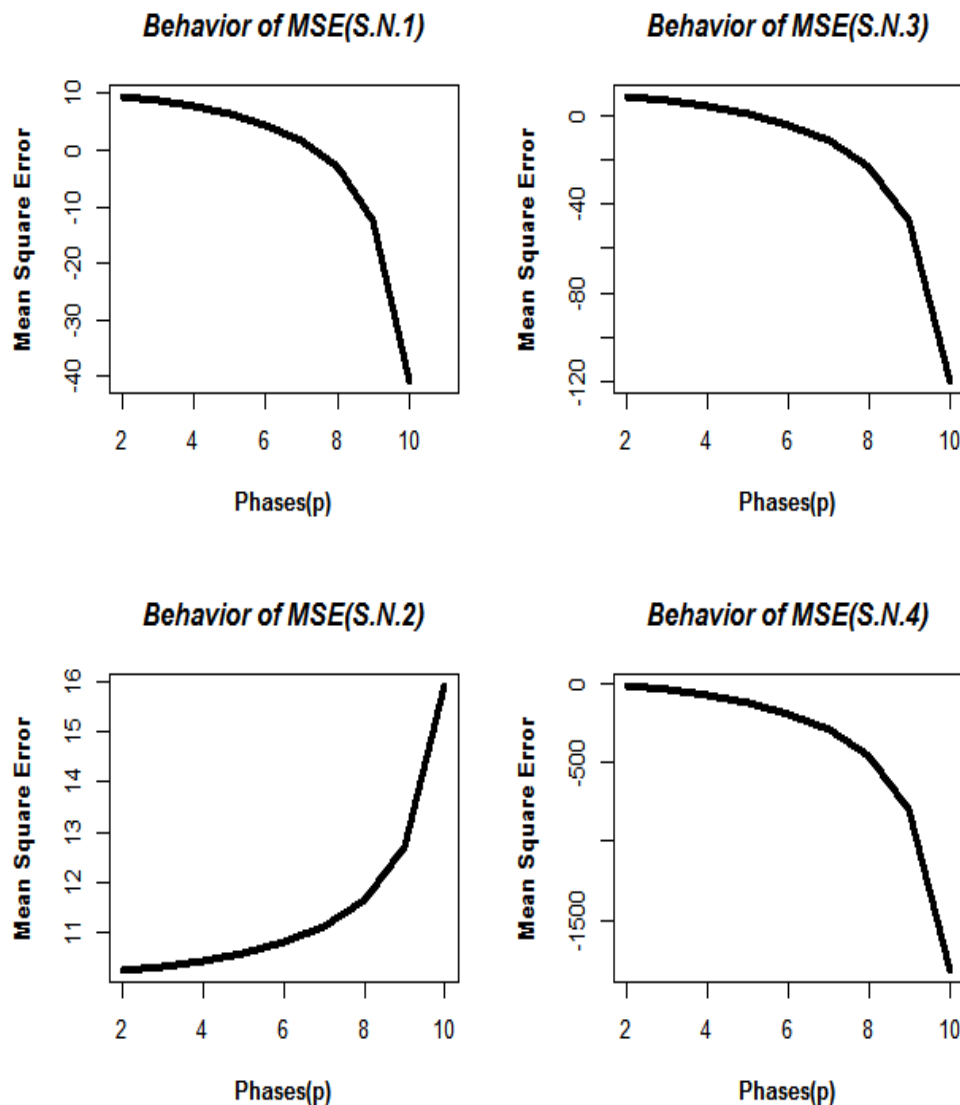
We have consider four different combinations on the basic of degree of correlation between auxiliary variables. Each of high positive and high negative correlation between understudy and auxiliary variables is combined with corresponding high and low correlation between auxiliaries. One the ground of four cases we computed MSE of proposed estimator which are presented in Table 5.2.5.

From the Table 5.2.5 we can observe for combinations of correlation at serial no 1 in Table 7.2.4, the MSE has shown decreasing pattern with the increase in the phase. Up to phase 7 MSE has become very low. Similarly, for combinations of serial no 3 the MSE has shown decline via increases in phases. Comparatively, for Serial no 3 the decrease in MSE is rapid than the decrease in MSE of serial no 3. For, serial number 2 the MSE almost remains constant intra-phase. After each phase there is a very minor increase MSE. For serial number 4 MSE starts from negative values. Again it is not suspicious as we have used population quantities to find MSE. If we wise to make comparison we can use magnitude of MSE in such case.

Table 5.2.5: $MSE(\bar{y}_{F_{2(p)}})$ at Combinations of High Correlation

Sample Size	Phase	$MSE(\bar{y}_{F_{2(p)}})$			
		S.N#1	S.N#2	S.N#3	S.N#4
200	2	9.53	10.26	8.56	-12.42
180	3	8.74	10.33	6.54	-40.67
160	4	7.73	10.44	3.95	-76.98
140	5	6.37	10.59	0.5	-125.4
120	6	4.48	10.81	-4.32	-193.1
100	7	1.64	11.13	-11.5	-294.8
80	8	-3.09	11.66	-23.6	-464.3
60	9	-12.5	12.72	-47.7	-803.2
40	10	-41	15.92	-120.2	-1820.1

The pattern of $MSE(\bar{y}_{F_{6(p)}})$ for all serial numbers 1 to 4 is presented graphically in figure 7.2.2. We observed that for serial number 1 and 3 MSE is showing decreasing pattern while for serial number 2 we have upward trend but it is not so sharp. By observing the numerical results and graphical pattern of MSE we can conclude that the proposed estimator will give better performance with the increasing phase when there is combination of correlations is like serial number 1, 3as mentioned in table 7.2.4

Figure 5.2.2: Behavior of $MSE(\bar{y}_{F_{2(p)}})$ at Combinations of High Correlation**Table 5.2.6: Relative Efficiency of $\bar{y}_{F_{2(p)}}$ at Combinations of High Correlations**

Phase	Each phase/2nd phase(%)				Each phase/Previous phase(%)			
	S.N 1	S.N 2	S.N 3	S.N 4	S.N 1	S.N 2	S.N 3	S.N 4
3	91.71	100.68	76.40	327.46	91.71	100.68	76.40	327.46
4	81.11	101.75	46.14	619.81	88.44	101.06	60.40	189.28
5	66.84	103.22	5.84	1009.66	82.41	101.44	12.66	162.90
6	47.01	105.36	-50.47	1554.75	70.33	102.08	-864.0	153.99
7	17.21	108.48	-134.35	2373.59	36.61	102.96	266.20	152.67
8	-32.42	113.65	-275.70	3738.33	-188.41	104.76	205.22	157.50
9	-131.16	123.98	-557.24	6466.99	404.53	109.09	202.12	172.99
10	-430.22	155.17	-1404.2	14654.59	328.00	125.16	251.99	226.61

Table 5.2.6 presents measures of relative efficiency for combinations of high correlations presents in table 5.2.6. For combinations at serial number 1 and 3 both the measures of efficiencies are less than 100 up to phase five and phase four respectively and afterwards they becomes negative. This means that performance of proposed estimator will get better and better with increase in number of phase where there is combinations of correlations at serial number 1 and 3 in table 5.2.4. For serial number 2 the measures are boarding little above 100. It can be interpreted that MSE will not differ to large extend with increasing number of phases.

For serial number 4 since the MSE starts with negative values and kept on increasing in terms of magnitude so we conclude from this evidence that the combinations at serial number 4 in table 5.2.4 are not suitable for our proposed estimator.

As a final comment we say that in case we have combinations of high correlations in such a way that high positive or high negative correlation between understudy and auxiliaries combined with low positive or negative correlations between auxiliaries and our proposed estimator will produce better results with increasing phase.

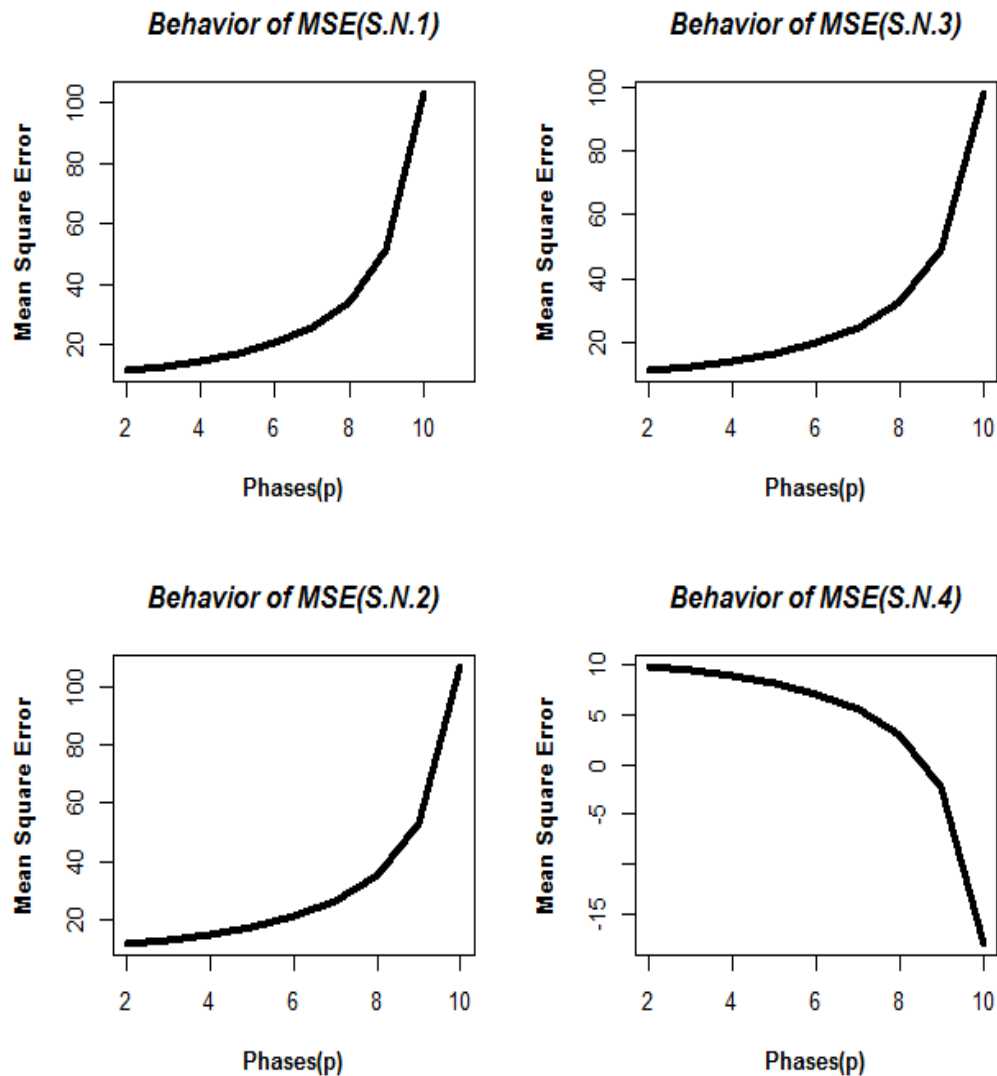
After discussing performance of proposed estimator (3.2) at combinations of high correlations. Now, we will examine its performance at low correlations. For this task consider table 5.2.7.

Table 5.2.7: Combinations of Low Levels of Correlation

S.N	Combinations of Correlation	\bar{Y}	ρ_{xy}	ρ_{xz}	ρ_{yz}	N
1	Low positive between understudy and auxiliaries Low positive between auxiliaries	48.0557	0.2406	0.20354	0.2547	1678
2	Low positive between understudy and auxiliaries high positive between auxiliaries	48.0557	0.2406	0.90354	0.2547	1678
3	Low negative between understudy and auxiliaries Low negative between auxiliaries	48.0557	-0.2406	-0.20354	-0.2547	1678
4	High negative between understudy and auxiliaries High negative between auxiliaries	48.0557	-0.2406	-0.90354	-0.2547	1678

Table 7.2.8: $MSE(\bar{y}_{F_{2(p)}})$ at Combinations of Low Correlation

Sample Size	Phase	$MSE(\bar{y}_{F_{2(p)}})$			
		S.N#1	S.N#2	S.N#3	S.N#4
200	2	11.32	11.37	11.26	9.82
180	3	12.76	12.87	12.61	9.39
160	4	14.61	14.80	14.36	8.83
140	5	17.08	17.36	16.68	8.08
120	6	20.54	20.96	19.94	7.04
100	7	25.72	26.36	24.82	5.47
80	8	34.36	35.35	32.96	2.87
60	9	51.64	53.33	49.24	-2.35
40	10	103.49	107.29	98.08	-18.01

Figure 5.2.3: Behavior of $MSE(\bar{y}_{F_{2(p)}})$ at Combinations of Low Correlation

For combinations of low level correlations presented in table 5.2.7 the results of MSE of proposed estimator (3.2) are presented in table 5.2.8. The corresponding graphical representation is in figure 5.2.3. From table 5.2.8 we observed that MSE of the estimator tends to increase for first three combinations of low level correlations. The incremental rate in all first three case is almost same. Also there is not a rapid uplift in MSE for all three cases. For fourth combination of correlations at low levels, MSE has the decreasing trend. The MSE shows continuous decrease up to phase 8.

The graphical pattern of MSE of the results produced in table 5.2.8 shows upward slopes of first three combinations. The slope for fourth combination is downwards which testify that MSE has a decreasing trend.

Table 5.2.9 is demonstration of relative efficiency according to all four combinations in Table 5.2.7. We observed that for first three combination both type of efficiencies are more than hundred. Which means that with increasing phase the relative efficiency of the estimator decline. Whereas, for the fourth combination the relative efficiencies are less than 100. In fact with the increase in phases the MSE rapidly decline and performance of the estimator gets better and better.

Table 5.2.9: Relative Efficiency of $\bar{y}_{F_{2(p)}}$ at Combinations of Low Correlations

Phase	Each phase/2nd phase(%)				Each phase/Previous phase(%)			
	S.N 1	S.N 2	S.N 3	S.N 4	S.N 1	S.N 2	S.N 3	S.N 4
3	112.72	113.18	112.05	95.57	112.72	113.18	112.05	95.57
4	129.07	130.13	127.55	89.88	114.51	114.97	113.83	94.04
5	150.87	152.73	148.21	82.29	116.89	117.37	116.20	91.56
6	181.40	184.36	177.14	71.67	120.23	120.71	119.52	87.09
7	227.19	231.82	220.53	55.73	125.24	125.74	124.49	77.76
8	303.50	310.91	292.84	29.17	133.59	134.12	132.79	52.34
9	456.12	469.10	437.47	-23.96	150.29	150.88	149.39	-82.13
10	913.99	943.65	871.36	-183.33	200.38	201.16	199.18	765.26

6. Conclusions and Recommendations

On the grounds of mathematical results, mathematical comparisons, constructed families and empirical studies of proposed estimators (3.1) and (3.2) we can draw following conclusions

1. Our proposed Estimators are generalized p-phased which provide flexibility to go up to any phase of sampling. Furthermore, for every desired phase we do not have to construct mathematical expressions right from the word go. We have readymade expressions of MSE and just need to replace desired value of p .

2. The proposed estimator-I is generalized p -phased and estimators by Hanif et al (2015) are now special cases of the proposed estimators-I, for $p = 3$ and $p = 4$ respectively.
3. The proposed estimator-II is also generalized p -phased and estimators by Samiuddin and Hanif (2007), Hanif et al (2015) and proposed estimator-I are now special cases of the proposed estimators-II, for different conditions over p, α_i and β_i .
4. Based upon the results of empirical study conducted for proposed estimator-I we can conclude that the MSE of the estimator will have increasing trend with increasing phases. This conclusion is based upon generalized results in contrast to the same conclusion drawn by Hanif et al (2015), who just utilized third and fourth phase.
5. In case of large population we observed that MSE of (3.1) are very close to each other. Therefore, the loss in efficiency will be negligible if we wish to go beyond 2nd phase. In this way we can get maximum information out of samples as well as the desired principal of repetition can also be achieved under NIC.
6. The empirical study for proposed estimator-II revealed that for all possible eight combinations of correlations between variables only two types of results of MSE are produced. This is because of the structural formulation of the expression of MSE of (3.2) presented in (5.1.27).
7. For moderate-low correlation in both cases the behavior of $MSE(\bar{y}_{F2})$ is anti. For all positive case $MSE(\bar{y}_{F2})$ tends to increase with number of phases. For 2nd case $MSE(\bar{y}_{F2})$ has decreasing pattern. Hence, we can conclude that estimator (3.2) will be useful under the situation of case (ii). It will not only reduce MSE but also the efficiency of the estimation will also be enhanced.
8. For other different combinations of correlations we may conclude that estimator (3.2) will perform better by reducing MSE and increasing efficiency when there is (i)-High positive correlation between y and x , y and z and low positive between z and x . (ii)-High positive correlation between y and x , y and z and low negative between z and x . (iii)-High negative correlation between y and x , y and z and low positive between z and x . In all other case there is a smooth steady and slow increase in MSE per phase.

References

1. Adhvaryu, D., Gupta, P.C. (1983). On some alternative sampling strategies using auxiliary information. *Metrika*, 30, 4, 217-226.
2. Chand, L. (1975). Some ratio type estimators based on two or more auxiliary variables. Unpublished Ph.D. thesis, Iowa State University, Ames, Iowa (USA).
3. Gupta, P. C. (1978). On some quadratic and higher degree ratio and product estimators.
4. Hanif, M., Hamad, N. and Shahbaz, M. Q. (2010) Some New Regression Types Estimators in Two Phase Sampling, *World App. Sci. J.*, Vol. 8(7), 799–803.

5. Hanif .S, Butt. N. S and Shahbaz. M. Q. (2015). On estimation in three and four phase sampling. *Sci.Int (Lahore)*, 27(3), 2575-2578.
6. Khare, B.B. and Srivastava, S.R. (1981). A general regression ratio estimator for the population mean using two auxiliary variables. *Aligarh Journal of Statistics*, 1, 43-51.
7. Kiregyera, B. (1984). A Regression-Type Estimator using two Auxiliary Variables and Model of Double sampling from Finite Populations. *Metrika*. 31, 215-226.
8. Mohanty, S. (1967). Combination of Regression and Ratio Estimate. *Jour. Ind. Statist. Asso.*, 5,16-19.
9. Sahai, A. (1979). An efficient variant of the product and ratio estimator. *Statist. Neerlandica*, 33, 27-35.
10. Samiuddin, M. and Hanif, M. (2006). Estimation in two phase sampling with complete and incomplete information. *Proc. 8th Islamic Countries Conference on Statistical Sciences*. 13, 479-495.
11. Sahoo, J. and Sahoo, L.N. (1993). A class of estimators in two-phase sampling using two auxiliary variables. *J. Ind. Soc. Agri. Statist.*, 31, 107-114.
12. Sahoo, J. and Sahoo, L.N. (1994). On the efficiency of four chain-type estimators in two phase sampling under a model. *Statistics*, 25, 361-366.
13. Singh, H.P and Tailor. R,” Use of correlation coefficient in estimating the finite population mean”, *Statistics in Transition*.Vol.6(4),Pp.555-560,2003
14. Sisodia, B.V.S and Dwivedi, V.K “A modified ratio estimator using coefficient of variation of auxiliary variable”, *Jour.Ind.Soc.Agri.Stat.*, Vol.33(1), Pp.13-18, 1981.
15. Srivastava, S.K. (1966). On ratio and linear regression method of estimation with several auxiliary variables. *J. Ind. Statist. Assoc.*, 4, 66-72.
16. Srivastava, S.K. (1967). An estimator using auxiliary information in sample surveys. *Calcutta Statistical Association Bulletin*, 16, 121-132.
17. Vos, J.W.E. (1980). Mixing of direct ratio and product method estimators. *Statist. Neerlandica*, 34, 209-218.
18. Walsh, J. E. (1970). Generalization of ratio estimate for population total. *Sankhya*, (A), 32, č. 1, s. 99 – 106.