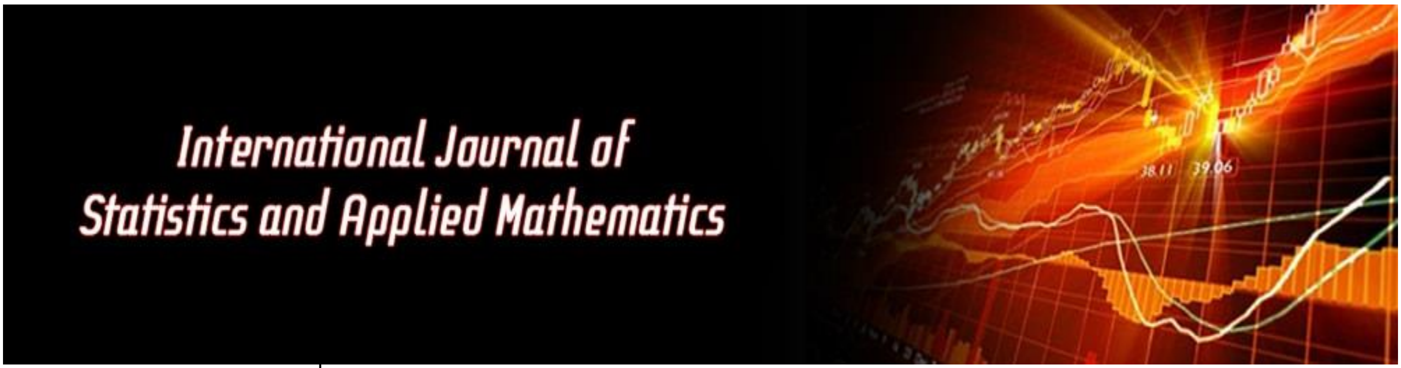

ISSN: 2456-1452
Maths 2024; 9(3): 27-33

https://www.mathsjournal.com
Received: 10-03-2024
Accepted: 12-04-2024

Naveen GP
Department of Mathematics and Statistics CCSHAU, Hisar, Haryana, India

Manoj Kumar
Department of Mathematics and Statistics CCSHAU, Hisar, Haryana, India

Corresponding Author:
Naveen GP
Department of Mathematics and Statistics CCSHAU, Hisar, Haryana, India

Alternate estimator of population mean under compromise method of imputation

Naveen GP and Manoj Kumar

Abstract
The analysis of data is complicated when two or more values are missing. Missing data generates bias and makes it more difficult to handle and analysis of data, which reduces efficiency. Imputation is a method where by imputed values are used to fill the missing values based on the data that is already existed and available axillary information. There are various techniques available to impute the values. After imputing the missing observations, we obtained a complete data set that can be analysed using some traditional methods. Previously imputation procedures like Mean method of Imputation, Ratio method of imputation, compromised method of imputation procedures are used in order to impute the missing observations. In this study an alternate estimator of population mean using estimator given by Singh *et al*. (2016) under compromised imputation method is given. The expression for bias and mean squared error (MSE) up to the first order approximation are derived. An numerical study is also carried out in order to compare the efficiency of the estimator with the previously existing estimators.

Keywords: Population mean, auxiliary information, mean method of imputation, ratio method of imputation, and compromise imputation

Introduction
Nonresponses or missing values are common occurrences in sample surveys. Item nonresponse and unit nonresponse are the two types of nonresponse. Item nonresponse is deal with the imputation method, while unit nonresponse is deal with the weight method. Similarly, the most common method for solving missing values is by imputation, which replaces missing values using the existing data as a source. Additionally, many researchers have studied the auxiliary information available in order to increase the accuracy of population mean estimation simple random sampling without replacement (SRSWOR). For example, Cochran applied the auxiliary information at the estimation stage and proposed an estimator to estimate the population mean. Bahl and Tuteja (1991) [2] first proposed new ratio-type exponential method for estimating the mean of population using information on auxiliary variable and they are more efficient than the existing estimators.

Let the mean of the population $\pi = \{1,2,\ldots,N\}$ is $\bar{Y} = \frac{1}{N}\sum_{i=1}^{n} y_i$. A sample s of size n is drawn from the population by simple random sampling without replacement (SRSWOR) to estimate population mean $(\bar{Y})$. Let the number of responding units be r out of sampled units. The set of responding units be R and the remaining non-responding units be R^c. For every units i ∈ R , the value of y_i is observed and the remaining units i ∈ R^c are missed and have to be imputed using the imputation techniques. Here assume that the imputation has carried using the available auxiliary information x such that x_i, the values x for unit i. is known and it is positive for all i ∈ s. The following notations used are are given by Lee *et al*. (1994) for the case of single value imputation.

If the ith unit requires the imputation, the value of $\hat{b}x_i$ is imputed where $= \frac{\sum_{i\in R} y_i}{\sum_{i\in R} x_i}$. The data after the imputation becomes

$$y_i = \begin{cases} y_i \ if \ i \in R \\ \hat{b}x_i \ if \ i \in R^c \end{cases} \tag{1}$$

The above imputation method used is known as Ratio method of imputation. The point estimator of population mean under this method of imputation is given by

$$\bar{y}_{RAT} = \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r}$$

Where

$$\bar{x}_n = \frac{1}{n}\sum_{i\in s} x_i$$

$$\bar{x}_r = \frac{1}{r}\sum_{i\in R} x_i$$

$$\bar{y}_r = \frac{1}{r}\sum_{i\in R} y_i$$

Another method of imputation is mean method of imputation. In this method the data after imputation becomes

$$y_i = \begin{cases} y_i \; if \; i \in R \\ \bar{y}_r \; if \; i \in R^c \end{cases} \tag{2}$$

The point estimator for the above method of imputation is given by

$$\bar{y}_M = \frac{1}{r}\sum_{i\in R} y_i = \bar{y}_r$$

Singh and Horn (2000) [15] have another imputation procedure called compromise imputation which is giving more efficient result than the other existing imputation procedures. The data after imputation becomes

$$y_i = \begin{cases} \alpha \frac{n}{r} y_i + (1-\alpha)\hat{b}x_i \; if \; i \in R \\ (1-\alpha)\hat{b}x_i \; if \; i \in R^c \end{cases} \tag{3}$$

The point estimator of the population mean is given by

$$\bar{y}_{COMP} = \alpha\bar{y}_r + (1-\alpha)\bar{y}_r \frac{\bar{x}_n}{\bar{x}_r}$$

Singh and Deo (2003) [10] introduced an innovative power transformation estimator for approximating the population mean. When applying. their imputation approach, the data exhibit the following format:

$$y_{.i} = \left\{ \begin{matrix} y_i \;\; if \; i\epsilon R \\ \bar{y}_r \left[n \left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\alpha} - r \right] \frac{x_i}{\sum_{i\epsilon R^c}^{n} x_i} \;\; if \; \epsilon R^c \end{matrix} \right\} \tag{4}$$

Where α is an appropriately. Chosen constant, so that the resultant estimator variance is minimum. The point estimator (4) for above imputation technique becomes:

$$\bar{y}_{SD} = \bar{y}_r (\frac{\bar{x}_n}{\bar{x}_r})^{\alpha}$$

If α = 0 then $\bar{y}_{SD} = \bar{y}_r$ and if α = 1 then $\bar{y}_{SD} = \bar{y}_{RAT}$, $\alpha = \rho_{yx}\frac{C_y}{C_x}$

1. Theory

Let us define

$$\varepsilon = \frac{\bar{y}_r}{\bar{Y}} - 1$$

$$\delta = \frac{\bar{x}_r}{\bar{X}} - 1$$

$$\eta = \frac{\bar{x}_n}{\bar{X}} - 1$$

Rao and Sitter (1995) gave the concept of two-phase sampling and the mechanism of that for given r and n we have

$E(\varepsilon) = E(\delta) = E(\eta) = 0$

and

$E(\varepsilon^2) = (\frac{1}{r} - \frac{1}{N})C^2{}_y$, $E(\delta^2) = \left(\frac{1}{r} - \frac{1}{N}\right) C^2{}_x$, $E(\eta^2) = \left(\frac{1}{n} - \frac{1}{N}\right) C^2{}_x$

$E(\varepsilon\delta) = (\frac{1}{r} - \frac{1}{N})\rho C_y C_x$, $E(\delta\eta) = (\frac{1}{n} - \frac{1}{N})C^2{}_x$, $E(\varepsilon\eta) = (\frac{1}{n} - \frac{1}{N})\rho C_y C_x$

Where,

$$C^2{}_y = \frac{S^2{}_y}{\bar{Y}^2}$$

$$C^2{}_x = \frac{S^2{}_x}{\bar{X}^2}$$

$$\rho = \frac{S_{xy}}{(S_x S_y)}$$

The properties (Bias and MSE) of the above discussed imputation methods are as follows,

1. The bias and MSE of Ratio method of imputation is,

- $B(\bar{y}_{RAT}) \approx (\frac{1}{r} - \frac{1}{n})\bar{Y}(C^2{}_x - \rho C_y C_x)$ (5)
- $MSE(\bar{y}_{RAT}) \approx \left(\frac{1}{n} - \frac{1}{N}\right) S^2{}_y + (\frac{1}{r} - \frac{1}{n})[S^2{}_y + R^2{}_1 S^2{}_x - 2R_1 S_{xy}]$ (6)

2. The bias and MSE of the compromised method of imputation is,

- $B(\bar{y}_{COMP}) \approx (1-\alpha)(\frac{1}{r} - \frac{1}{n})\bar{Y}(C^2{}_x - \rho C_y C_x)$ (7)
- $MSE(\bar{y}_{COMP}) \approx MSE(\bar{y}_{RAT}) - (\frac{1}{r} - \frac{1}{n})\left(1 - \rho\frac{C_y}{C_x}\right)^2 \bar{Y}^2 C^2{}_x$ (8)

Where,

$\alpha = 1 - \rho\frac{C_y}{C_x}$ (Suitable constant to reduce the variance)

3. The bias and MSE of the Singh and Deo (2003) [10] method of imputation is

- $B(\bar{y}_{SD}) = \left(\frac{1}{r} - \frac{1}{N}\right)\bar{Y}\left[\frac{\alpha(\alpha-1)}{2} C_x{}^2 - \alpha\rho C_y C_x\right]$ (9)
- $Min.MSE\ (\bar{y}_{SD}) = MSE(\bar{y}_{rat}) - (\frac{1}{r} - \frac{1}{n})S^2{}_x\left(\rho S_X S_Y - \frac{\bar{Y}}{\bar{X}}\right)^2$ (10)

4. Proposed imputation method for missing data

By using the estimator given by Singh *et al*. (2016) [11], here we proposed a new estimator for missing data. The data after imputation takes as,

$$Y_i = \begin{cases} k\frac{n}{r}y_i + (1-k)\bar{y}_r d \ if\ i \in R \\ (1-k)\bar{y}_r d\ if\ i \in R^c \end{cases} \quad (11)$$

Where,

$$d = \bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right) exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)$$

$\bar{y}_r = \frac{\sum_{i=1}^{r} y_i}{r}$ and $\bar{x}_r = \frac{\sum_{i=1}^{r} x_i}{r}$

$$k = 1 - \frac{2[\rho C_y C_x - C_y{}^2]}{C_x{}^2}$$

Theorem 3.1

The point estimator of population mean ($\bar{Y}$) under proposed method imputation is

$$\bar{y}_p = k\bar{y}_r + (1-k)\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right) \quad (12)$$

Proof: we have

$$\bar{y}_p = \frac{1}{n}\sum_{i\in s} y_i$$

$$= \frac{1}{n}\left[\sum_{i\in R} y_i + \sum_{i\in R^c} y_i\right]$$

$$= \frac{1}{n}\left[\sum_{i\in R}\left[k\frac{n}{r}y_i + (1-k)\bar{y}_r d\right] + \sum_{i\in R^c}(1-k)\bar{y}_r d\right]$$

$$= \frac{1}{n}\left[k\frac{n}{r}r\bar{y}_r + (1-k)dr\bar{y}_r + (1-k)d(n-r)\bar{y}_r\right]$$

$$= \frac{1}{n}k\frac{n}{r}r\bar{y}_r + \frac{(1-k)}{n}d[r\bar{y}_r + (n-r)\bar{y}_r]$$

$$= k\bar{y}_r + (1-k)d\,\bar{y}_r$$

Then the point estimator is given as

$$\bar{y}_p = k\bar{y}_r + (1-k)\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)$$

Theorem 3.2
The bias of the proposed point estimator is given as

$$B(\bar{y}_p) = \bar{Y}(1-k)\left(\frac{1}{r}-\frac{1}{N}\right)C_x\left[\frac{3}{4}C_x - \frac{1}{2}\rho C_y\right] + \left(\frac{1}{n}-\frac{1}{N}\right)C_x\left[\rho C_y + \frac{1}{2}\right]{C_x}^2 \quad (13)$$

Proof: see appendix 1.

Theorem 3.3
The mean squared error (MSE) of the proposed estimator $\bar{y}_p$ is given by

$$MSE(\bar{y}_p) = \bar{Y}^2\left(\frac{1}{r}-\frac{1}{N}\right)\left[{C_y}^2 + (1-k)^2\frac{{C_x}^2}{4} - (1-k)\rho C_y C_x\right] \quad (14)$$

Proof: See appendix 2.

5. Efficiency comparison of the proposed estimator

In the following section the proposed estimator $(\bar{\boldsymbol{y}}_{\boldsymbol{p}})$ efficiency is compared with previous existing estimators i.e., ratio and compromised estimator by using MSE and an estimator with preferred smaller value of MSE.

6. Comparison of proposed estimator with Ratio estimator

From the expressions (14) and (6), we have $MSE(\bar{y}_p) < MSE(\bar{y}_{RAT})$ when

$$\bar{Y}^2\left(\frac{1}{r}-\frac{1}{N}\right)\left[{C_y}^2 + (1-k)^2\frac{{C_x}^2}{4} - (1-k)\rho C_y C_x\right] - \left(\frac{1}{n}-\frac{1}{N}\right){S_y}^2 + \left(\frac{1}{r}-\frac{1}{n}\right)\left(R^2{S_x}^2 - 2RS_{xy}\right) < 0 \quad (15)$$

Is true for optimum value. Hence therefore proposed estimator is more efficient than the Ratio estimator.

7. Comparison of proposed estimator with compromised imputation estimator

Form the expressions (14) and (8), we have $MSE(\bar{\boldsymbol{y}}_{\boldsymbol{p}})$ < Min. MSE $(\bar{y}_{Com})$ when

$$\left(\frac{1}{r}-\frac{1}{n}\right)\left[(1-\alpha)^2C^2{}_x - 2(1-\alpha)\rho C_y C_x\right] - \left(\frac{1}{r}-\frac{1}{N}\right)\left[(1-k)^2\frac{{C_x}^2}{4} - (1-k)\rho C_y C_x\right] < 0 \quad (16)$$

Is true for optimum value of $= 1 - \frac{\rho C_y}{\theta C_x}$, hence the proposed estimator is more efficient than the existing estimator.

8. Comparison of proposed estimator with Singh and Deo (2003) [10] estimator

Form the expressions (14) and (10), we have $MSE(\bar{\boldsymbol{y}}_{\boldsymbol{p}}) <$ Min. MSE $(\bar{y}_{SD})$ when

$$\bar{Y}^2\left(\frac{1}{r}-\frac{1}{N}\right)\left[(1-k)^2\frac{{C_x}^2}{4}-(1-k)\rho C_y C_x\right]-\bar{Y}^2\left[\alpha^2\left\{\left(\frac{1}{n}-\frac{1}{N}\right){C_x}^2+\left(\frac{1}{r}-\frac{1}{N}\right){C_x}^2-2\left(\frac{1}{n}-\frac{1}{N}\right)\rho C_x C_y\right\}+2\alpha\left\{\left(\frac{1}{n}-\frac{1}{N}\right)\rho C_x C_y-\left(\frac{1}{r}-\frac{1}{N}\right)\rho C_x C_y\right\}\right]<0 \quad (17)$$

Which is always true. Thus, the proposed estimator is more efficient than the Singh and Deo (2003) [10] estimator under optimality condition.

9. Case study

In this section, we assess the performance of the proposed estimator by conducting a comparative analysis against existing estimators. To facilitate this comparison, we employ two distinct population datasets. Here we computed the approximate value MSE of the respective estimators. The population 1 is taken from the NSSO and another population 2 was given by Kadilar and Cingi (2008). The performance of the estimators is studied based on the percentage relative efficiency (PRE).
The percentage relative efficiency is given by as follows:

$$PRE=\frac{V(\bar{y}_r)}{MSE(\bar{y}_i)}*100$$

Where,
i = $\bar{y}_{rat}$, $\bar{y}_{comp}$ and $\bar{y}_{prapo}$

Table 1: Data used for study

Parameters	Population 1(NSSO)	Population 2
N	2397	19
n	500	10
r (assumed)	350	8
$\bar{Y}$	2.121819	575
$\bar{X}$	2.170371	13537.68
C_Y	0.94899	1.4928
C_X	1.886679	0.956248
$\beta_1(X)$	4.754219	-
$\beta_2(X)$	33.11113	-
ρ_{YX}	-0.03895	0.88

In the following table the percentage relative efficiency of the different estimator are given which is used for comparison of the study. According to the following table it is clearly showing that the proposed estimator is performing superior than the existing estimators (Ratio and compromised method of imputation).

Table 2: MSE of estimators

Estimators	Population 1	Population 2
$\bar{y}_{rat}$	17294.42	132.6477
$\bar{y}_{comp}$	42162.14	136.2262
$\bar{y}_{prapo}$	42372.48	690.5383

The table 2 showing the PRE of Proposed estimator and the existing estimators (Ratio estimator and compromise imputation estimator). Form the above table for both the populations PRE of proposed estimator is more than the existing estimators by this we can conclude that the proposed estimator is preferable over the existing estimators by obtaining efficient point estimator.

10. Conclusion

A problem of non-response for a specific unit or units in the population occurs frequently in surveys involving the medical profession, social surveys, and household surveys etc… Due to missing values in the data set, this non-sampling error may creep and produce wrong inferences. In the present study, the alternative estimator of population mean under compromised imputation method has been proposed. Its bias and MSE also been proposed. For the proposed estimator the efficiency has been checked and prove to be more efficient than the existing estimators like ratio method of imputation, mean method of imputation and compromised imputation.

References

1. Adhvaryu D, Gupta PC. On some alternative sampling strategies using auxiliary information. Metrika. 1983;30(1):217-226.
2. Bahl S, Tuteja R. Ratio and product type exponential estimators. J Inf Optim Sci. 1991;12(1):159-164.
3. Chodjuntug K, Lawson N. A chain ratio exponential-type compromised imputation for mean estimation: case study on ozone pollution in Saraburi, Thailand. J Probab Stat. 2020.
4. Gira AA. Estimation of population mean with a new imputation method. Appl Math Sci. 2015;9(34):1663-1672.

5. Khoshnevisan M, Singh R, Chauhan P, Sawan N. A general family of estimators for estimating population mean using known value of some population parameter(s). Infinite Study. 2007.
6. Murthy MN. Sampling theory and methods. Sampling theory and methods. 1967.
7. Prasad S. A study on new methods of ratio exponential type imputation in sample surveys. Hacet J Math Stat. 2018;47(5):1281-1301.
8. Singh AK, Singh P, Singh VK. Exponential-type compromised imputation in survey sampling. J Stat Appl Prob. 2014;3(2):211.
9. Singh BK, Gogoi U. Estimation of Population mean using ratio-cum-product imputation techniques in sample survey. Res Rev: J Stat. 2018;7:38-49.
10. Singh GN. On the improvement of product method of estimation in sample surveys. J Indian Soc Agric Stat. 2003;56(3):267-265.
11. Singh GN, Maurya S, Khetan M, Kadilar C. Some imputation methods for missing data in sample surveys. Hacet J Math Stat. 2016;45(6):1865-1880.
12. Singh HP, Pal SK. A new chain ratio-ratio-type exponential estimator using auxiliary information in sample surveys. rn. 2015;55:7.
13. Singh S, Horn S. Compromised imputation in survey sampling. Metrika. 2000;51(3):267-276.

Appendix – 1: To prove 3.2 theorem, we have the point estimator as

$$\bar{y}_p = k\bar{y}_r + (1-k)\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)$$

expressing in terms of ε, δ *and* η

$$\bar{y}_p = k\bar{y}_r + (1-k)\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)exp\left(\frac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right)$$

$$= k\bar{Y}(1-\varepsilon) + (1-k)\bar{Y}(1-\varepsilon)\left(\frac{\bar{X}(1+\eta)}{\bar{X}(1+\delta)}\right)exp\left(\frac{\bar{X}-\bar{X}(1+\delta)}{\bar{X}+\bar{X}(1+\delta)}\right)$$

$$= k\bar{Y}(1-\varepsilon) + (1-k)\bar{Y}(1-\varepsilon)\left(\frac{(1+\eta)}{(1+\delta)}\right)exp\left(\frac{\bar{X}[1-(1+\delta)]}{\bar{X}[1-(1+\delta)]}\right)$$

$$= k\bar{Y}(1-\varepsilon) + (1-k)\bar{Y}(1-\varepsilon)\left(\frac{1+\eta}{1+\delta}\right)exp\left(\frac{\bar{X}[1-(1+\delta)]}{\bar{X}[1-(1+\delta)]}\right)$$

$$= k\bar{Y}(1-\varepsilon) + (1-k)\bar{Y}(1-\varepsilon)\left(\frac{1+\eta}{1+\delta}\right)exp\left[\frac{-\delta}{2+\delta}\right]$$

$$= k\bar{Y}(1-\varepsilon) + (1-k)\bar{Y}(1-\varepsilon)\left(\frac{1+\eta}{1+\delta}\right)exp\left[\frac{-\delta}{2}\left(1-\frac{\delta}{2}\right)^{-1}\right]$$

$$= k\bar{Y}(1-\varepsilon) + (1-k)\left(\frac{1+\eta}{1+\delta}\right)\bar{Y}\left[1+\varepsilon-\frac{\delta}{2}+\frac{\delta^2}{4}+\frac{\epsilon\delta}{2}\right]$$

$$= k\bar{Y}(1-\varepsilon) + (1-k)\bar{Y}(1+\eta)(1+\delta)^{-1}\left[1+\varepsilon-\frac{\delta}{2}+\frac{\delta^2}{4}+\frac{\epsilon\delta}{2}\right]$$

$$= k\bar{Y}(1-\varepsilon) + (1-k)\bar{Y}(1+\eta)(1-\delta+\delta^2\ldots)\left[1+\varepsilon-\frac{\delta}{2}+\frac{\delta^2}{4}+\frac{\epsilon\delta}{2}\right]$$

$$= k\bar{Y}(1-\varepsilon) + (1-k)\bar{Y}(1-\delta+\delta^2+\eta-\eta\delta+\eta\delta^2)\left[1+\varepsilon-\frac{\delta}{2}+\frac{\delta^2}{4}+\frac{\epsilon\delta}{2}\right]$$

$$= k\bar{Y}(1-\varepsilon) + (1-k)\bar{Y}\left[1+\varepsilon-\frac{\delta}{2}+\frac{\delta^2}{4}+\frac{\epsilon\delta}{2}-\varepsilon\delta+\frac{\delta^2}{2}+\eta\varepsilon-\frac{\eta\delta}{2}\right]$$

$$= \bar{Y}+\bar{Y}\varepsilon + (1-k)\bar{Y}\left[-\frac{\delta}{2}+\frac{\delta^2}{4}-\frac{\epsilon\delta}{2}+\frac{\delta^2}{2}+\eta\varepsilon-\frac{\eta\delta}{2}\right]$$

$$\bar{y}_p = \bar{Y} + \bar{Y}\varepsilon + (1-k)\bar{Y}\left[-\frac{\delta}{2} + \frac{\delta^2}{4} - \frac{\epsilon\delta}{2} + \frac{\delta^2}{2} + \eta\varepsilon - \frac{\eta\delta}{2}\right]$$

Taking expectation on both side

$$\bar{y}_p = \bar{Y} + \bar{Y}\varepsilon + (1-k)\bar{Y}\left[-\frac{\delta}{2} + \frac{\delta^2}{4} - \frac{\epsilon\delta}{2} + \frac{\delta^2}{2} + \eta\varepsilon - \frac{\eta\delta}{2}\right]$$

$$E(\bar{y}_p) - \bar{Y} = +\bar{Y}E(\varepsilon) + (1-k)\bar{Y}\left[-\frac{1}{2}E(\delta) + \frac{3}{4}E(\delta^2) - \frac{1}{2}E(\epsilon\delta) + E(\eta\varepsilon) - \frac{1}{2}E(\eta\delta)\right]$$

$$B(\bar{y}_p) = \bar{Y}(1-k)\left[\frac{3}{4}\left(\frac{1}{r} - \frac{1}{N}\right){C_x}^2 - \frac{1}{2}\left(\frac{1}{r} - \frac{1}{N}\right)\rho C_y C_x + \left(\frac{1}{n} - \frac{1}{N}\right)\rho C_y C_x + \frac{1}{2}\left(\frac{1}{n} - \frac{1}{N}\right){C_x}^2\right]$$

$$B(\bar{y}_p) = \bar{Y}(1-k)\left(\frac{1}{r} - \frac{1}{N}\right)C_x\left[\frac{3}{4}C_x - \frac{1}{2}\rho C_y\right] + \left(\frac{1}{n} - \frac{1}{N}\right)C_x\left[\rho C_y + \frac{1}{2}\right]{C_x}^2$$

Appendix – 2: To prove 3.3 theorem, we have

$$\left(\bar{y}_p - \bar{Y}\right)^2 = \bar{Y}^2\varepsilon^2 + \bar{Y}^2(1-k)^2\left[\frac{\delta^2}{4}\right] + 2(1-k)\bar{Y}^2\left(\frac{-\varepsilon\delta}{2}\right)$$

Taking expectation on both side

$$E\left(\bar{y}_p - \bar{Y}\right)^2 = \bar{Y}^2E(\varepsilon^2) + \bar{Y}^2(1-k)^2\left[\frac{1}{4}E(\delta^2)\right] - (1-k)\bar{Y}^2E(\varepsilon\delta)$$

$$MSE(\bar{y}_p) = \bar{Y}^2\left(\frac{1}{r} - \frac{1}{N}\right){C_y}^2 + \bar{Y}^2(1-k)^2\left[\frac{1}{4}\left(\frac{1}{r} - \frac{1}{N}\right){C_x}^2\right] - (1-k)\bar{Y}^2\left(\frac{1}{r} - \frac{1}{N}\right)\rho C_y C_x$$

$$MSE(\bar{y}_p) = \bar{Y}^2\left(\frac{1}{r} - \frac{1}{N}\right)\left[{C_y}^2 + (1-k)^2\frac{{C_x}^2}{4} - (1-k)\rho C_y C_x\right]$$

Differentiating w.r.t (1-k) and equating to zero

$$\frac{d[MSE(\bar{y}_p)]}{d[1-k]} = \bar{Y}^2\left(\frac{1}{r} - \frac{1}{N}\right)\left[{C_y}^2 + 2(1-k)\frac{{C_x}^2}{4} - \rho C_y C_x\right] = 0$$

$${C_y}^2 + 2(1-k)\frac{{C_x}^2}{4} - \rho C_y C_x = 0$$

$$(1-k)\frac{{C_x}^2}{2} = \rho C_y C_x - {C_y}^2$$

$$(1-k) = \frac{2[\rho C_y C_x - {C_y}^2]}{{C_x}^2}$$

$$k = 1 - \frac{2[\rho C_y C_x - {C_y}^2]}{{C_x}^2}$$