4.1 Introduction

In survey sampling, the estimation of population variance of the variable under study has also attracted the attention of large number of statisticians. It is well known that the use of the auxiliary information can increase the efficiency of the estimators of parameters of interest. Using the prior knowledge of population variance $\sigma^2_x$ of auxiliary variable x, which is highly correlated with study variable y, several estimators have been defined by different authors such as Das and Tripathi (1978), Srivastava and Jhajj (1980), Ahmed et al. (2003), Jhajj et al. (2005), Kadilar and Cingi (2006), Pradhan (2010) in the literature for estimating the unknown population variance of study variable y. In chapter 3, we have explained that when information on population parameters of auxiliary variable x is not known in advance then generally two phase (double) sampling design has been widely used in which value of unknown population parameters of auxiliary...
variable $x$ such as population variance $\sigma_x^2$ is estimated on the basis of first phase sample and is used for construction of estimator of parameter of interest based on second phase sample.

In the present chapter, we propose a generalized difference-cum-ratio type estimator for the population variance under double sampling design. The expressions for bias and mean square error of the proposed estimator have been obtained. The comparison of the proposed estimator has been made with the linear regression type estimator and sample variance. Effort has also been made to illustrate the results numerically and graphically.

## 4.2 Notations and Results

A preliminary large random sample (first phase sample) of size $n'$ is drawn from a finite population of size $N$ and both auxiliary variable $x$ and study variable $y$ are measured on it. Then second phase random sample of size $n (< n')$ is drawn from the first phase sample.

Let $Y_i$ and $X_i$ denote the respective values of variables $y$ and $x$ on the $i^{th}$ ($i = 1, 2, ..., N$) unit of the population and the corresponding small letters denote the values in the sample.
Denoting

\[
\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i
\]

\[
\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i
\]

\[
\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i,
\]

\[
\bar{y}' = \frac{1}{n'} \sum_{i=1}^{n'} y_i'
\]

\[
s^2_y = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2,
\]

\[
s'^2_y = \frac{1}{n'-1} \sum_{i=1}^{n'} (y_i' - \bar{y}')^2
\]

\[
S^2_y = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2
\]

\[
S'^2_y = \frac{1}{N'-1} \sum_{i=1}^{N'} (Y_i' - \bar{Y}')^2
\]

\[
s'^2_x = \frac{1}{n'-1} \sum_{i=1}^{n'} (x_i' - \bar{x})^2,
\]

\[
s'^2_x = \frac{1}{n'-1} \sum_{i=1}^{n'} (x_i' - \bar{x})^2
\]

\[
\mu_{rs} = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \bar{Y})' (X_i - \bar{X})^s
\]

\[
\lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{r/2} \mu_{02}^{s/2}}
\]

\[
\rho_v = \frac{(\lambda_{22} - 1)}{\sqrt{(\lambda_{04} - 1)(\lambda_{40} - 1)}
\]

where \(s'^2_x\), \(s'^2_y\) and \(s'^2_y\), \(s'^2_y\) are the sample variances of variables x and y based on the sampling units of first and second phase samples of sizes \(n'\) and \(n\) respectively.
Defining

\[ \delta_0 = \frac{s_x^2}{S_x^2} - 1 \]
\[ \epsilon_0 = \frac{s_y^2}{S_y^2} - 1 \]
\[ \delta_1 = \frac{s_x^2}{S_x^2} - 1 \]
\[ \epsilon_1 = \frac{s_y^2}{S_y^2} - 1 \]

We assume that

\[ E(\delta_0) = E(\delta_1) = E(\epsilon_0) = E(\epsilon_1) = 0 \]

\[ E(\delta_0^2) = \frac{Var(s_x^2)}{S_x^4} \]
\[ E(\epsilon_0^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\epsilon_0\epsilon_1) = \frac{Cov(s_x^2, s_y^2)}{S_y^4} \]

\[ E(\delta_0\epsilon_0) = \frac{Cov(s_x^2, s_x^2)}{S_x^2 S_y^2} \]
\[ E(\delta_1\epsilon_0) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\delta_0\delta_1) = \frac{Cov(s_x^2, s_x^2)}{S_x^4} \]
\[ E(\delta_1\epsilon_1) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\epsilon_0\epsilon_1) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\epsilon_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\delta_1^2) = \frac{Var(s_x^2)}{S_x^4} \]

\[ E(\delta_0^2) = \frac{Var(s_x^2)}{S_x^4} \]

\[ E(\delta_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\epsilon_0^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\epsilon_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\delta_0\delta_1) = \frac{Cov(s_x^2, s_x^2)}{S_x^4} \]

\[ E(\delta_0\epsilon_1) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\delta_1\epsilon_0) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\epsilon_0\epsilon_1) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\epsilon_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\epsilon_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\delta_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\epsilon_0^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\epsilon_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\delta_0\delta_1) = \frac{Cov(s_x^2, s_x^2)}{S_x^4} \]

\[ E(\delta_0\epsilon_1) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\delta_1\epsilon_0) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\epsilon_0\epsilon_1) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\epsilon_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\epsilon_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\delta_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\epsilon_0^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\epsilon_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\delta_0\delta_1) = \frac{Cov(s_x^2, s_x^2)}{S_x^4} \]

\[ E(\delta_0\epsilon_1) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\delta_1\epsilon_0) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\epsilon_0\epsilon_1) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\epsilon_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\epsilon_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\delta_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\epsilon_0^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\epsilon_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\delta_0\delta_1) = \frac{Cov(s_x^2, s_x^2)}{S_x^4} \]

\[ E(\delta_0\epsilon_1) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\delta_1\epsilon_0) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\epsilon_0\epsilon_1) = \frac{Cov(s_x^2, s_y^2)}{S_x^2 S_y^2} \]

\[ E(\epsilon_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\epsilon_1^2) = \frac{Var(s_y^2)}{S_y^4} \]

\[ E(\delta_1^2) = \frac{Var(s_y^2)}{S_y^4} \]
For the sake of justification of the above results obtained, we derive the following result:

\[
E(\delta_1 \varepsilon_1) = E\left(\left(\frac{s_{x}^2}{S_x^2} - 1\right)\left(\frac{s_{y}^2}{S_y^2} - 1\right)\right)
\]

\[
= \frac{1}{S_x^2 S_y^2} E\left[(s_{x}^2 - S_x^2)(s_{y}^2 - S_y^2)\right]
\]

\[
= \frac{\text{Cov}(s_{x}^2, s_{y}^2)}{S_x^2 S_y^2}
\]

On the similar line, other results involved in (4.2.1) can be derived.

### 4.3 The Proposed Estimator

Under the sampling design defined in section (4.2), we propose an estimator of population variance \(S_y^2\) of study variable y as

\[
\hat{S}_{hgd}^2 = \left[ s_y^2 + \theta \left( s_{y}^2 - s_y^2 \right) \right] \left[ \frac{s_{x}^2}{s_x^2 + \theta \left( s_{y}^2 - s_x^2 \right)} \right]^\alpha
\]  \hspace{1cm} (4.3.1)

where \(\alpha\) and \(\theta\) are unknown constants.
To find the bias and mean square error of estimator \( \hat{s}_{hgd}^2 \), up to first order of approximation, we expand \( \hat{s}_{hgd}^2 \) in terms of \( \varepsilon' s \) and \( \delta' s \) and retaining terms up to second degree of approximation, as

\[
\hat{s}_{hgd}^2 = S_y^2 \left[ 1 + \theta_0 + \theta (\varepsilon_1 - \varepsilon_0) + \alpha(1-\theta)(\delta_1 - \delta_0) - \alpha \delta_0 (\delta_1 - \delta_0) + \alpha \theta^2 (\delta_1 - \delta_0)^2 \\
+ 2\alpha \theta \delta_0 (\delta_1 - \delta_0) - \alpha \theta \delta_1 (\delta_1 - \delta_0) + \frac{\alpha(\alpha-1)}{2}(1-\theta)^2 (\delta_1 - \delta_0)^2 \\
+ \alpha(1-\theta)\varepsilon_0 (\delta_1 - \delta_0) + \alpha \theta (1-\theta)(\varepsilon_1 - \varepsilon_0)(\delta_1 - \delta_0) \right] 
\]  

(4.3.2)

Taking expectation of (4.3.2) and using the results of section (4.2), we obtain

\[
E(\hat{s}_{hgd}^2) = S_y^2 + S_y^2 (1-\theta)^2 \left\{ \frac{\alpha(\alpha+1)}{2} \left( \frac{V(s_x^2) - V(s_x'^2)}{S_x^4} \right) - \alpha \left( \frac{Cov(s_y^2, s_x^2) - Cov(s_y^2, s_x'^2)}{S_y^2 S_x^2} \right) \right\} 
\]

\[
\Rightarrow \text{Bias}(\hat{s}_{hgd}^2) = S_y^2 (1-\theta)^2 \left\{ \frac{\alpha(\alpha+1)}{2} \left( \frac{V(s_x^2) - V(s_x'^2)}{S_x^4} \right) - \alpha \left( \frac{Cov(s_y^2, s_x^2) - Cov(s_y^2, s_x'^2)}{S_y^2 S_x^2} \right) \right\} 
\]

(4.3.3)
From (4.3.2), the mean square error (MSE) of the estimator $\hat{s}_{hgd}^2$ is obtained by using the results of section (4.2), up to first order of approximation, as

$$MSE\left(\hat{s}_{hgd}^2\right) = E\left(\hat{s}_{hgd}^2 - s_y^2\right)^2$$

$$= S_y^4 E\left[\varepsilon_0 + \theta (\varepsilon_1 - \varepsilon_0) + \alpha (1-\theta) (\delta_1 - \delta_0)\right]^2$$

$$= V\left(s_y^2\right) + \left(\theta^2 - 2\theta\right)\left\{ V\left(s_y^2\right) - V\left(s_x^{t2}\right) \right\}$$

$$+ \alpha^2 (1-\theta)^2 \frac{S_y^4}{S_x^4}\left\{ V\left(s_x^{t2}\right) - V\left(s_x^{t2}\right) \right\}$$

$$- 2\alpha (1-\theta)^2 \frac{S_y^2}{S_x^2}\left\{ Cov\left(s_y^2,s_x^2\right) - Cov\left(s_y^{t2},s_x^{t2}\right) \right\}$$

(4.3.4)

The expression (4.3.4) depends upon two unknown constants $\alpha$ and $\theta$. So keeping the value of $\theta$ fixed, we differentiate (4.3.4) w.r.t. $\alpha$ and equating to zero, we get

$$2\alpha (1-\theta)^2 \frac{S_y^4}{S_x^4}\left\{ V\left(s_x^{t2}\right) - V\left(s_x^{t2}\right) \right\} - 2 (1-\theta)^2 \frac{S_y^2}{S_x^2}\left\{ Cov\left(s_y^2,s_x^2\right) - Cov\left(s_y^{t2},s_x^{t2}\right) \right\} = 0$$

77
Chapter-4: A Generalized Difference-Cum-Ratio Type Estimator for the Population Variance in Double Sampling

After solving, we get the optimum value of $\alpha$ as

$$
(\alpha)_{opt} = \frac{S_x^2 \left\{ \text{Cov} \left( s_y^2, s_x^2 \right) - \text{Cov} \left( s_y^2, s_x^{r2} \right) \right\}}{S_y^2 \left\{ V \left( s_x^2 \right) - V \left( s_x^{r2} \right) \right\}} \tag{4.3.5}
$$

Substituting the optimum value of $\alpha$ from (4.3.5) in (4.3.4), we get minimum mean square error, $\text{Min.MSE} \left( \hat{s}_{hgd}^2 \right)$, as

$$
\text{Min.MSE} \left( \hat{s}_{hgd}^2 \right) = V \left( s_y^2 \right) + \left( \theta^2 - 2\theta \right) \left\{ V \left( s_y^2 \right) - V \left( s_y^{r2} \right) \right\} - \left( 1 - \theta \right)^2 \frac{\left\{ \text{Cov} \left( s_y^2, s_x^2 \right) - \text{Cov} \left( s_y^2, s_x^{r2} \right) \right\}^2}{\left\{ V \left( s_x^2 \right) - V \left( s_x^{r2} \right) \right\}} \tag{4.3.6}
$$

**Theorem 4.3.1:** Up to first order of approximation, the bias of estimator $\hat{s}_{hgd}^2$ is

$$
\Rightarrow \text{Bias} \left( \hat{s}_{hgd}^2 \right) = S_y^2 \left( 1 - \theta \right)^2 \left\{ \frac{\alpha (\alpha + 1)}{2} \left( \frac{V \left( s_x^2 \right) - V \left( s_x^{r2} \right)}{S_x^4} \right) - \alpha \left( \frac{\text{Cov} \left( s_y^2, s_x^2 \right) - \text{Cov} \left( s_y^2, s_x^{r2} \right)}{S_y^2 S_x} \right) \right\}
$$

78
and its Mean Square Error is given by

\[
MSE(\hat{s}_{hgd}^2) = V\left(s^2_y\right) + \left(\theta^2 - 2\theta\right)\left(V\left(s^2_y\right) - V\left(s^2_x\right)\right) + \alpha^2 \left(1 - \theta\right)^2 \frac{S^4_y}{S^4_x} \left(V\left(s^2_x\right) - V\left(s^2_x\right)\right)
- 2\alpha \left(1 - \theta\right)^2 \frac{S^2_y}{S^2_x} \left(Cov\left(s^2_y, s^2_x\right) - Cov\left(s^2_y, s^2_x\right)\right)
\]

Theorem 4.3.2: Up to first order of approximation, the MSE of \(\hat{s}_{hgd}^2\) is minimized for

\[
(\alpha)_{opt} = \frac{S^2_x \left\{Cov\left(s^2_y, s^2_x\right) - Cov\left(s^2_y, s^2_x\right)\right\}}{S^2_y \left(V\left(s^2_x\right) - V\left(s^2_x\right)\right)}
\]

and its minimum value is given by

\[
Min.MSE(\hat{s}_{hgd}^2) = V\left(s^2_y\right) + \left(\theta^2 - 2\theta\right)\left(V\left(s^2_y\right) - V\left(s^2_x\right)\right)
- \left(1 - \theta\right)^2 \frac{\left\{Cov\left(s^2_y, s^2_x\right) - Cov\left(s^2_y, s^2_x\right)\right\}^2}{\left(V\left(s^2_x\right) - V\left(s^2_x\right)\right)}
\]

The results given in theorem 4.3.1 and theorem 4.3.2 are for any random sampling design which satisfies the conditions of section (4.2). For comparing the
proposed estimator with the existing ones, we hereby obtain the results under simple random sampling technique used for selection of samples as a special case.

**Special Case:** Using simple random sampling procedure for selection of samples in given double sampling design, we have

\[
\begin{align*}
Var(s_y^2) &= \frac{1}{n} S_y^4 (\lambda_{40} - 1) \\
Var(s_x^2) &= \frac{1}{n} S_x^4 (\lambda_{04} - 1) \\
Cov(s_x^2, s_y^2) &= \frac{1}{n} S_x^2 S_y^2 (\lambda_{22} - 1)
\end{align*}
\]

For the sake of justification of the above results obtained, we derive the following result:

\[
Cov(s_x^2, s_y^2) = E \left[ \left( s_x^2 - S_x^2 \right) \left( s_y^2 - S_y^2 \right) \right]
\]

\[
= E \left( s_x^2 s_y^2 \right) - S_x^2 S_y^2
\]

\[
= E \left[ \left( \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right) \left( \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \right) \right] - S_x^2 S_y^2
\]
\[
E \left[ \sum_{i=1}^{n} (x_i - \bar{X})^2 \sum_{i=1}^{n} (y_i - \bar{Y})^2 \right] = \frac{(N-n)n}{(N-1)} \mu_{22} + \frac{nN(n-1)}{(N-1)} \mu_{20} \mu_{02}
\]

(4.3.9)

\[
E \left[ (\bar{X} - \bar{X})^2 \sum_{i=1}^{n} (y_i - \bar{Y})^2 \right] = \frac{(N-n)(N-2n)}{(N-1)(N-2)n} \mu_{22} + \frac{N(N-n)(n-1)}{n(N-1)(N-2)} \mu_{20} \mu_{02}
\]

(4.3.10)

\[
E \left[ (\bar{Y} - \bar{Y})^2 \sum_{i=1}^{n} (x_i - \bar{X})^2 \right] = \frac{(N-n)(N-2n)}{(N-1)(N-2)n} \mu_{22} + \frac{N(N-n)(n-1)}{n(N-1)(N-2)} \mu_{20} \mu_{02}
\]

(4.3.11)
Chapter-4: A Generalized Difference-Cum-Ratio Type Estimator for the Population Variance in Double Sampling

\[ E \left( \frac{(\bar{X} - \bar{X})^2}{(\bar{Y} - \bar{Y})^2} \right) = \frac{\left( N - n \right) \left[ N^2 - (6n - 1)N + 6n^2 \right]}{(N - 1)(N - 2)(N - 3)n^3} \mu_{22} \]
\[ + \frac{N \left( N - n \right) \left( N - n - 1 \right) (n - 1)}{(N - 1)(N - 2)(N - 3)n^3} \mu_{20}\mu_{02} \]
\[ + \frac{2N \left( N - n \right) \left( N - n - 1 \right) (n - 1)}{(N - 1)(N - 2)(N - 3)n^3} \mu_{11}^2 \]

(4.3.12)

Using the results of (4.3.9), (4.3.10), (4.3.11) and (4.3.12) in (4.3.8), and retaining terms up to first order of approximation, we have

\[ \text{Cov}(s^2_x, s^2_y) = \frac{1}{n} S^2_x S^2_y (\lambda_{22} - 1) \]

On the similar line, other results involved in (4.3.7) can be derived.

Substituting results of (4.3.7) in (4.3.3) and (4.3.4) respectively, we have

\[ \text{Bias}(\hat{s}^2_{hgd}) = \left\{ \frac{1}{n} - \frac{1}{n'} \right\} S^2_y (1 - \theta)^2 \left\{ \frac{\alpha (\alpha + 1)}{2} (\lambda_{04} - 1) - \alpha (\lambda_{22} - 1) \right\} \]

(4.3.13)

\[ \text{MSE}(\hat{s}^2_{hgd}) = \frac{1}{n} S^4_y (\lambda^2_{40} - 1) + \left\{ \frac{1}{n} - \frac{1}{n'} \right\} S^4_y \left\{ (\theta^2 - 2\theta) (\lambda_{40} - 1) \right. \]
\[ + \alpha^2 (1 - \theta)^2 (\lambda_{04} - 1) - 2\alpha (1 - \theta)^2 (\lambda_{22} - 1) \left\} \]

(4.3.14)
For minimizing $MSE(\hat{s}_{hgd}^2)$, we differentiate (4.3.14) w.r.t. $\alpha$ and equating to zero

$$2\alpha (1-\theta)^2 (\lambda_{04} - 1) - 2(1-\theta)^2 (\lambda_{22} - 1) = 0$$

After some simplification of above equation, we get

$$(\alpha)_{opt} = \frac{(\lambda_{22} - 1)}{(\lambda_{04} - 1)}$$  (4.3.15)

Substituting the optimum value of $\alpha$ from (4.3.15) in (4.3.14), we obtain

$$Min.MSE(\hat{s}_{hgd}^2) = \frac{1}{n} S_y^4 (\lambda_{40} - 1) + \left(\frac{1}{n} - \frac{1}{n'}\right) S_y^4 (\lambda_{40} - 1) \left[ (\theta^2 + 2\theta) - (1-\theta)^2 \rho_v^2 \right]$$  (4.3.16)

From (4.3.16), we see that $Min.MSE(\hat{s}_{hgd}^2)$ will vary with the change in the value of $\theta$. So the range of variation of $\theta$ has been obtained in the next section (4.4) at which the proposed estimator is better than the existing ones.

**Cor 4.3.1.1:** Up to first order of approximation, under double sampling
design in which simple random sampling is used at both phases, the bias of estimator $\hat{s}^2_{hgd}$ is

$$\text{Bias}\left(\hat{s}^2_{hgd}\right) = \left(\frac{1}{n} - \frac{1}{n'}\right)S_y^2\left(1-\theta\right)^2 \left\{\frac{\alpha(\alpha+1)}{2}\left(\lambda_{04} - 1\right) - \alpha\left(\lambda_{22} - 1\right)\right\}$$

and its MSE is

$$\text{MSE}\left(\hat{s}^2_{hgd}\right) = \frac{1}{n}S_y^4\left(\lambda_{40} - 1\right) + \left(\frac{1}{n} - \frac{1}{n'}\right)S_y^4\left\{\left(\theta^2 - 2\theta\right)(\lambda_{40} - 1) + \alpha^2\left(1-\theta\right)^2\left(\lambda_{04} - 1\right) - 2\alpha\left(1-\theta\right)^2\left(\lambda_{22} - 1\right)\right\}$$

Cor 4.3.2.1: Up to first order of approximation, under double sampling design in which simple random sampling is used at both phases, the MSE of $\hat{s}^2_{hgd}$ is minimized for

$$\left(\alpha\right)_{opt} = \frac{\left(\lambda_{22} - 1\right)}{\left(\lambda_{04} - 1\right)}$$

and its minimum value is given by
Chapter 4: A Generalized Difference-Cum-Ratio Type Estimator for the Population Variance in Double Sampling

\[
\text{Min.MSE}\left(\hat{s}_{\text{hgd}}^2\right) = \frac{1}{n} S_y^4 (\lambda_{40} - 1) + \left(\frac{1}{n} - \frac{1}{n'}\right) S_y^4 (\lambda_{40} - 1) \left[\left(\theta^2 - 2\theta\right) - (1 - \theta)^2 \rho_v^2\right]
\]

### 4.4 Comparison

For comparing the proposed estimator with the sample variance \(s_y^2\) and linear regression type estimator, \(s_{\text{ird}}^2 = s_y^2 + \frac{\text{Cov}(s_y^2, s_v^2)}{\text{Var}(s_y^2)} (s_v^2 - s_y^2)\), we first obtain the expressions of their mean square errors, under sampling design defined in section (4.2) using simple random sampling at both stages, up to first order of approximation, as

\[
\text{MSE}\left(\hat{s}_{\text{ird}}^2\right) = \frac{1}{n} S_y^4 (\lambda_{40} - 1) - \left(\frac{1}{n} - \frac{1}{n'}\right) S_y^4 (\lambda_{40} - 1) \rho_v^2 \quad (4.4.1)
\]

\[
\text{MSE}\left(s_y^2\right) = \frac{1}{n} S_y^4 (\lambda_{40} - 1) \quad (4.4.2)
\]

Using (4.3.16) and (4.4.1) and after some algebra, we obtain

\[
\text{MSE}\left(s_{\text{ird}}^2\right) - \text{Min.MSE}\left(\hat{s}_{\text{hgd}}^2\right) = \left(\frac{1}{n} - \frac{1}{n'}\right) S_y^4 (\lambda_{40} - 1) \left(2\theta - \theta^2\right)(1 - \rho_v^2) \quad (4.4.3)
\]
On the R.H.S. of (4.4.3), the factors \( \left( \frac{1}{n} - \frac{1}{n'} \right), S^4_y, (\lambda_{40} - 1) \) and 
\( (1 - \rho^2_v) \) are always positive. So their product 
\( \left( \frac{1}{n} - \frac{1}{n'} \right) S^4_y (\lambda_{40} - 1)(1 - \rho^2_v) \) is also positive. Hence R.H.S. of (4.4.3) will be positive for 
\( 0 < \theta < 2 \).

\[
\Rightarrow \text{MSE} \left( s^2_{brd} \right) - \text{Min.MSE} \left( \hat{s}^2_{hgd} \right) > 0 \quad \text{if} \quad 0 < \theta < 2
\]

or \( \text{Min.MSE} \left( \hat{s}^2_{hgd} \right) < \text{MSE} \left( s^2_{brd} \right) \quad \text{if} \quad 0 < \theta < 2 \)  

(4.4.4)

Similarly, using (4.3.16) and (4.4.2), we obtain

\[
\text{MSE} \left( s^2_y \right) - \text{Min.MSE} \left( \hat{s}^2_{hgd} \right) = \left( \frac{1}{n} - \frac{1}{n'} \right) S^4_y (\lambda_{40} - 1) \left[ \rho^2_v - (\theta^2 - 2\theta)(1 - \rho^2_v) \right]
\]

\[
\Rightarrow \text{MSE} \left( s^2_y \right) - \text{Min.MSE} \left( \hat{s}^2_{hgd} \right) > 0 \quad \text{if} \quad 1 - \frac{1}{\sqrt{1 - \rho^2_v}} < \theta < 1 + \frac{1}{\sqrt{1 - \rho^2_v}}
\]

or \( \text{Min.MSE} \left( \hat{s}^2_{hgd} \right) < \text{MSE} \left( s^2_y \right) \quad \text{if} \quad 1 - \frac{1}{\sqrt{1 - \rho^2_v}} < \theta < 1 + \frac{1}{\sqrt{1 - \rho^2_v}} \)  

(4.4.5)
4.5 Numerical Illustration

To have a rough idea about the gain in efficiency of the proposed estimator \( \hat{s}_{hgd}^2 \) over the sample variance \( s_y^2 \) and linear regression type estimator \( s_{lrd}^2 \) under double sampling, we take the empirical population considered in the literature (Source: Sukhatme & Sukhatme, 1997, p-256). The values of the population parameters obtained are given in table 4.5.1. The mean square error and relative efficiency of the proposed estimator \( \hat{s}_{hgd}^2 \) w.r.t. linear regression type estimator \( s_{lrd}^2 \) and sample variance \( s_y^2 \) under the same sampling design are given for some different values of \( \theta \) in the table 4.5.2.

Table 4.5.1: Values of Population Parameters

<table>
<thead>
<tr>
<th>( N )</th>
<th>( n' )</th>
<th>( n )</th>
<th>( S_y )</th>
<th>( \rho_v )</th>
<th>( \lambda_{40} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>89</td>
<td>45</td>
<td>23</td>
<td>716.65</td>
<td>0.81</td>
<td>3.81</td>
</tr>
</tbody>
</table>
Table 4.5.2: Mean Square Errors and Relative Efficiency

(a) Proposed Estimator vs. Regression-type Estimator

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$MSE\left( s_{trd}^2 \right)$</th>
<th>$MSE\left( \hat{s}_{hgd}^2 \right)$</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$s_{trd}^2$</td>
<td>$\hat{s}_{hgd}^2$</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>56329.12</td>
<td>56329.12</td>
<td>148.30</td>
</tr>
<tr>
<td>0.5</td>
<td>56329.12</td>
<td>46105.85</td>
<td>148.30</td>
</tr>
<tr>
<td>1</td>
<td>56329.12</td>
<td>42698.09</td>
<td>148.30</td>
</tr>
<tr>
<td>1.5</td>
<td>56329.12</td>
<td>46105.85</td>
<td>148.30</td>
</tr>
<tr>
<td>2</td>
<td>56329.12</td>
<td>56329.12</td>
<td>148.30</td>
</tr>
</tbody>
</table>
Figure 4.5.1: Graphical representation of Efficiency of Proposed estimator and Regression-type estimator.

\[ \rho_v = 0.81 \]
(b) Proposed Estimator vs. Sample variance

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$MSE\left(s_y^2\right)$</th>
<th>$MSE\left(\hat{s}_{hgd}^2\right)$</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$s_y^2$</td>
<td>$\hat{s}_{hgd}^2$</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>83539.758</td>
<td>56329.12</td>
<td>100</td>
</tr>
<tr>
<td>0.5</td>
<td>83539.758</td>
<td>46105.85</td>
<td>100</td>
</tr>
<tr>
<td>1</td>
<td>83539.758</td>
<td>42698.09</td>
<td>100</td>
</tr>
<tr>
<td>1.5</td>
<td>83539.758</td>
<td>46105.85</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>83539.758</td>
<td>56329.12</td>
<td>100</td>
</tr>
</tbody>
</table>
Figure 4.5.2: Graphical representation of Efficiency of Proposed Estimator and Sample Variance estimator.
From table 4.5.2, we can see that there is a significant gain in efficiency of the proposed estimator \( \hat{s}_{\text{hgd}}^2 \) over the linear regression type estimator \( s_{\text{lr}}^2 \) in double sampling for \( 0 < \theta < 2 \) and sample variance \( s_y^2 \) for \( 1 - \frac{1}{\sqrt{1 - \rho^2}} < \theta < 1 + \frac{1}{\sqrt{1 - \rho^2}} \). From the fig. 4.5.1 and fig. 4.5.2, we also see that for \( 0 < \theta < 2 \), the efficiency of proposed estimator is more than the regression-type estimator and it is also more efficient than sample variance for wider range of \( \theta \) at moderate value of correlation coefficient. Hence we conclude that proposed estimator will always be better than the existing regression-type estimator under double sampling for \( 0 < \theta < 2 \) and sample variance for \( 1 - \frac{1}{\sqrt{1 - \rho^2}} < \theta < 1 + \frac{1}{\sqrt{1 - \rho^2}} \).

### 4.6 Summary

When information on population parameters of auxiliary variable such as population mean \( \bar{X} \), population variance \( S_x^2 \) etc is not known in advance then two phase (double) sampling design has been widely used in which value of unknown population parameters of auxiliary variable \( x \) such as population variance \( S_x^2 \) is estimated on the basis of first phase sample and is used for
construction of estimator of parameter of interest based on second phase sample.

We propose a generalized difference-cum-ratio type estimator for the population variance under double sampling design. The expressions for bias and mean square error of the proposed estimator have been obtained under random sampling design. For comparing the proposed estimator with the regression type estimator and sample variance, the expressions for bias and mean square error have been obtained as a special case when simple random sampling technique is used for selection of samples. The range of variation of $\theta$ at which the proposed estimator is better than the existing ones have been obtained theoretically. The efficiency of the optimum estimator of the proposed estimator and that of the existing ones have also been shown numerically by taking a empirical population considered in the literature. The results have also been illustrated graphically. Hence we conclude that proposed estimator will always be better than the existing ones for a given range of variation of $\theta$ which can be made available from the past experience.