STATISTICS IN TRANSITION-new series, October 2011
Vol. 12, No. 2, pp. 357—380

IMPROVED ESTIMATORS OF COEFFICIENT OF VARIATION IN A FINITE POPULATION

V. Archana[1], K. Aruna Rao[2]

ABSTRACT

Coefficient of Variation (C.V) is a unitless measure of dispersion. Hence it is widely used in many scientific and social investigations. Although a lot of work has been done concerning C.V in the infinite population models, it has been neglected in the finite populations. Many areas of applications of C.V involves the finite populations like the use in official statistics and economic surveys of the World Bank. This has motivated us to propose six new estimators of the population C.V. In finite population studies regression estimators are widely used and the idea is exploited to propose the new estimators. Three of the proposed estimators are the regression estimators of the C.V for the study variable while the other three estimators makes use of the regression estimators of population mean and variance to estimate the ratio $\frac{\sigma_y}{\bar{Y}}$, the population C.V for the study variable. The bias and mean square error (MSE) of these estimators were derived for the simple random sampling design. The performance of these estimators is compared using two real life data sets. The simulation is carried out to compare the estimators in terms of coverage probability and the length of the confidence interval. The small sample comparison indicates that two of the proposed estimators perform better than the sample C.V. The regression estimator using the information on the Population C.V of the auxiliary variable emerges as the best estimator.

Key words: Model based comparison; Coefficient of Variation; Simple Random Sampling; Regression estimator; Mean Square Error; Confidence interval.

[1] Dept. of Statistics, Mangalore University, Mangalagangothri, Konaje, Karnataka,India. E-mail: archana_stkl@yahoo.com.

[2] Dept. of Statistics, Mangalore University, Mangalagangothri, Konaje, Karnataka,India. E-mail: arunaraomu@yahoo.com.

1. Introduction

During the last few years, the Coefficient of variation (C.V) has received the attention of many statisticians, although it was used as a measure of variation by scientists in other disciplines. The C.V is a relative measure of dispersion and is unitless. Thus, it facilitates the comparison of variability measured in different units. Although some investigators prefer the use of standard deviation to co-efficient of variation it is difficult in many instances to draw meaningful conclusions from standard deviation as it is an absolute measure. The coefficient of variation expressed in percentages indicates quickly the extent of variability present in the data. Some of the specific examples include the study of rainfall (Anantha Krishnan and Soman (1989), Business and Engineering (De, Ghosh and Wells (1996)). The C.V is also common in applied probability fields such as renewal theory, queuing theory and reliability theory. The C.V is also used in multiple time scales and the life time (Kordonsky and Gerts bakh.(1997)).

The research on C.V dates back to the work of McKay (1932), Pearson(1932), and Fieller(1932) where they have studied a numerical approximation to the distribution of the sample C.V (in the case of normality). Later on it was extended by Hendrick and Robey (1936) and Koopmans et al.(1964). Nairy and Rao (2003) and the references cited there discusses the various tests for testing the equality of C.V's of independent normal distributions. The research work on the C.V of the normal distribution is fast growing and one of the recent references is that of Mahmoudvand and Hassani (2007) who proposed two new confidence intervals for the C.V in a normal distribution.

Compared to the research work on C.V of the normal distribution, research on C.V of a finite population is of recent origin. The estimation of C.V in finite population was initially discussed by Das and Tripathi (1981a, b). Since then various researchers have attempted the estimation of C.V which include the works of Rajyaguru and Gupta (2002, 2006), Tripathi et al.(2002) , Patel and Shah (2009), among others. Following the idea of Srivastava (1971, 80) and Das and Tripathi (1980), Tripathi et al.(2002) constructed a general class of estimators of C.V. This class of estimators is a hybrid class in the sense that a regression type of estimators is used to construct a general class of ratio estimators of C.V. The ratio/product and regression estimators of C.V constructed from the sample C.V are members of this class. They also obtained an optimum estimator belonging to this class. In this paper we derive a general expression for the bias and MSE of the regression estimator of any parameter of interest θ_y using information on any parameter η_x of the auxiliary variable. The general expression for the MSE indicates that if we construct a regression estimator using any other estimator including a hybrid estimator of θ_y then the regression estimator thus obtained is more efficient than the hybrid estimator. The focal point of this paper is to compare the performance of seven regression/regression type estimators of C.V constructed from the sample C.V. In this comparison we have not included the optimum estimators of Tripathi et al.(2002). The reason for this is that, as indicated previously, we can always construct a more efficient regression estimator using this optimum estimator. Such estimators becomes complex in nature compared to the regression estimators based on sample C.V. Tripathi et al.(2002) compared the asymptotic performance of 22 estimators which includes the regression and regression type of estimators using two real life datasets. Patel and Shah (2009) compared the small sample MSE of five estimators of C.V which do not include the simple regression type of estimators based on sample C.V. In the last four decades a lot of papers have appeared on the regression estimators of other parameter of interest (like mean and variance). For some of these works see the references cited in Sahoo et al.(2003), Verma (2008) and Pradhan (2010). This has motivated us to undertake a comprehensive comparison of the regression/regression type of estimators constructed from sample C.V.

As a first step in this direction, we have proposed estimators of finite population C.V when the underlined sampling scheme is Simple Random Sample (with or without replacement). The first estimator is the sample C.V. Six new estimators are also proposed in the paper. Three of them are the regression estimators using the information on population C.V, mean and variance of an auxiliary variable. Other three estimators are ratio type regression estimators, where regression estimators are used for the estimation of population mean, and variance using information on the auxiliary variable. Bias and mean square error (MSE) of these estimators are derived to the order of $O(n^{-1})$. Since simulations based on a real life data setting cannot cover a wide variety of complexities regarding the performance of the estimators, we have resorted to model based comparison of the regression estimators. Using bivariate normal distribution the performance of the estimators is compared using i) small sample MSE, ii) coverage probability and iii) average length of the confidence interval. Extensive simulation is carried out covering a wide range of the correlation co-efficient between the study and auxiliary variable and various choice of the C.V of the auxiliary variable. Two of the six new estimators of the C.V perform better compared to the sample C.V. The regression estimator using the information on the C.V of the auxiliary variable emerges as the best estimator.

The organization of the paper is as follows. Section 2 presents the general expressions for bias and MSE of the regression estimators to the order of $O(n^{-1})$. The results are used to derive the bias and MSE of regression estimators of C.V constructed using sample C.V under simple random sampling. Comparisons of the asymptotic performance of the seven estimators are considered in section 3 using two real life data-sets. Section 4 deals with the small sample performance of these estimators. The final conclusions are presented in section 5.

2. General expression for the bias and MSE of the regression estimators.

Theorem 2.1: Let θ_y be the parameter of interest of the study variable 'y' to be estimated and let η_x denote a parameter of the auxiliary variable 'x'. Let $\hat{\theta}_y$ and $\hat{\eta}_x$ be their unbiased estimators then the regression estimator of θ_y is given by

$\hat{\theta}_{y_{\text{Re}g}} = \hat{\theta}_{\text{y}} + \hat{\beta}(\eta_{\text{x}} - \hat{\eta}_x)$, where β denotes the regression co-efficient of $\hat{\theta}_y$ on $\hat{\eta}_x$ and is given by,

$$\beta = \frac{\text{Cov}(\hat{\theta}_{\text{y}}, \hat{\eta}_x)}{V(\hat{\eta}_x)}$$

and $\hat{\beta}$ denotes an asymptotically unbiased estimator of β then the bias and MSE of the regression estimator $\hat{\theta}_{y_{\text{Re}g}}$ to the order of $O(\frac{1}{n})$ is given by

$$B(\hat{\theta}_{y_{\text{Re}g}}) = -\text{Cov}(\hat{\eta}_{\text{x}}, \hat{\beta}) + o\left(\frac{1}{n}\right) \tag{2.1}$$

and
$$M(\hat{\theta}_{y_{\text{Re}g}}) = \text{V}(\hat{\theta}_{\text{y}})(1-\rho^2) + o\left(\frac{1}{n}\right) \tag{2.2}$$

Proof: Using the ε - approach we have,
Let us take $\hat{\theta}_y = \theta_y(1+\varepsilon_1)$, $\hat{\eta}_{\text{x}} = \eta_x(1+\varepsilon_2)$ and $\hat{\beta} = \beta(1+\varepsilon_3)$
then

$$\begin{aligned}\hat{\theta}_{y_{\text{Re}g}} &= \hat{\theta}_y + \hat{\beta}(\eta_x - \hat{\eta}_x)\\ &= \theta_{\text{y}}(1+\varepsilon_1) + \beta(1+\varepsilon_3)(\eta_x - \eta_x(1+\varepsilon_2))\\ &= \theta_{\text{y}}(1+\varepsilon_1) - \beta\eta_x(1+\varepsilon_3)\varepsilon_2\end{aligned}$$

Now, if we take expectations on both sides we get

$$B(\hat{\theta}_{y_{\text{Re}g}}) = E(\hat{\theta}_{y_{\text{Re}g}} - \theta_y) = -\beta\eta_x E(\varepsilon_2\varepsilon_3) \qquad \because E(\varepsilon_1) = 0 \text{ and } E(\varepsilon_2) = 0$$

$$\begin{aligned}&= -\beta\eta_x \frac{Cov(\hat{\eta}_x, \hat{\beta})}{\beta\eta_x}\\ &= -Cov(\hat{\eta}_x, \hat{\beta})\end{aligned}$$

Similarly the variance of $\hat{\theta}_{y_{\mathrm{Re}\,g}}$ is given by

$$M(\hat{\theta}_{y_{\mathrm{Re}\,g}}) = E(\hat{\theta}_{y_{\mathrm{Re}\,g}} - \theta_y)^2 = \theta_y{}^2 \mathrm{V}(\varepsilon_1) + \beta^2\eta_x^2 \mathrm{V}(\varepsilon_2) - 2\theta_y\beta\eta_x Cov(\varepsilon_1, \varepsilon_2)$$

$$\because \ \mathrm{V}(\varepsilon_2\varepsilon_3) \text{ is of order } O\left(\frac{1}{n^2}\right).$$

$$= \theta_y{}^2 \frac{\mathrm{V}(\hat{\theta}_{\mathrm{y}})}{\theta_y{}^2} + \beta^2\eta_x^2 \frac{\mathrm{V}(\hat{\eta}_{\mathrm{x}})}{\eta_x^2} - 2\theta_y\beta\eta_x \frac{Cov(\hat{\theta}_y, \hat{\eta}_x)}{\theta_y\eta_x}$$

$$= \mathrm{V}(\hat{\theta}_y) - 2\beta Cov(\hat{\theta}_y, \hat{\eta}_x) + \beta^2 V(\hat{\eta}_x)$$

But $\beta = \dfrac{\mathrm{Cov}(\hat{\theta}_{\mathrm{y}}, \hat{\eta}_x)}{V(\hat{\eta}_x)}$. If we substitute β in the above equation we get

$$\mathrm{M}(\hat{\theta}_{\mathrm{y_{Reg}}}) = \mathrm{V}(\hat{\theta}_y) - \frac{2Cov^2(\hat{\theta}_y, \hat{\eta}_x)}{\mathrm{V}(\hat{\eta}_{\mathrm{x}})} + \frac{Cov^2(\hat{\theta}_y, \hat{\eta}_x)}{\mathrm{V}(\hat{\eta}_{\mathrm{x}})}$$

$$= \mathrm{V}(\hat{\theta}_y) - \frac{Cov^2(\hat{\theta}_y, \hat{\eta}_x)}{\mathrm{V}(\hat{\eta}_{\mathrm{x}})}$$

$$= \mathrm{V}(\hat{\theta}_{\mathrm{y}})(1 - \rho^2)$$

Hence, the proof follows.

Remark:

1. The expression for MSE to the order of $O(\frac{1}{n})$ does not change if we replace the unbiased estimators $\hat{\theta}_y$ and $\hat{\eta}_x$ by their asymptotically unbiased estimators of θ_y and η_x.

2. From the expression of MSE in (2.2), it becomes clear that if a regression estimator is constructed from an optimum estimator belonging to another class of estimators, this regression estimator is more efficient compared to the optimum estimator, although the decrease in the MSE may not be substantial.

2.2 SRSWR

In the sequel, we present the new estimators of C.V along with Bias and MSE.

2.2.1 Usual estimators $\left(\hat{\theta}_{y_1}\right)$:-

In theory of sampling it is customary to denote the study variable by ‘y’ and the auxiliary variable ‘x’. Let $\overline{Y}$ and $\sigma_y{}^2$ denote the population mean and population variance for the study variable.

In the following, $\sigma_y^{\,2}$ is also denoted σ_{yy} so as to generalize the notations for the higher order moments of the study and auxiliary variables. The primary focus of interest is to estimate $\theta_y = \dfrac{\sigma_y}{\overline{Y}}$. The usual estimator is obtained by using the sample mean and sample standard deviation as an estimators of the denominator and the numerator respectively. It is given by

$$\hat{\theta}_{y_1} = \frac{s_y}{\bar{y}}. \tag{2.3}$$

where $\bar{y}$ is the sample mean and $s_y^{\,2}$ is the sample variance and are given by

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \quad \text{and} \quad s_y^{\,2} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}.$$

Further, the bias and MSE of $\hat{\theta}_{y_1}$ are given by

$$bias\left(\hat{\theta}_{y_1}\right) = E\left(\hat{\theta}_{y_1} - \theta_{y_1}\right) = \left\{-\frac{1}{8n}\frac{\sigma_{yyyy}}{\overline{Y}\sigma_y^{\,3}} + \frac{1}{8n}\frac{(n-3)}{(n-1)}\frac{(\sigma_{yy})^2}{\overline{Y}\sigma_y^{\,3}} + \frac{1}{n}\left(\frac{\sigma_y}{\overline{Y}}\right)^3 - \frac{1}{2n}\frac{\sigma_{yyy}}{\sigma_y \overline{Y}^2}\right\} + O\left(\frac{1}{n^2}\right), \tag{2.4}$$

and

$$MSE\left(\hat{\theta}_{y_1}\right) = E\left(\hat{\theta}_{y_1} - \theta_{y_1}\right)^2 = \left\{\frac{1}{4n}\frac{\sigma_{yyyy}}{\sigma_y^{\,2}\overline{Y}^2} - \frac{(n-3)}{4n(n-1)}\frac{(\sigma_{yy})^2}{\sigma_y^{\,2}\overline{Y}^2} + \frac{1}{n}\left(\frac{\sigma_y}{\overline{Y}}\right)^4 - \frac{1}{n}\left(\frac{\sigma_{yyy}}{\overline{Y}^3}\right)\right\} + O\left(\frac{1}{n^2}\right). \tag{2.5}$$

2.2.2. Estimator $\left(\hat{\theta}_{y_2}\right)$

Let $\hat{\overline{Y}}_R$ denote the regression estimator of $\overline{Y}$. It is given by

$$\hat{\overline{Y}}_R = \bar{y} + b_1\left(\overline{X} - \bar{x}\right). \tag{2.6}$$

where b_1 is the estimator of B_1, the regression co-efficient of $\bar{y}$ on $\bar{x}$ and is given by,

$$B_1 = \frac{E(\bar{x}-\overline{X})(\bar{y}-\overline{Y})}{E(\bar{x}-\overline{X})^2} = \frac{\sigma_{xy}}{\sigma_{xx}}. \tag{2.7}$$

The estimator b_1 is obtained by substituting the corresponding sample moments in (2.6).

Using this regression estimator of $\overline{Y}$, the second estimator of $\hat{\theta}_{y_1}$ is given by

$$\hat{\theta}_{y_2} = \frac{s_y}{\bar{y} + b_1\left(\overline{X} - \bar{x}\right)}. \tag{2.8}$$

The bias and MSE of $\hat{\theta}_{y_2}$ are given by

$$bias\left(\hat{\theta}_{y_2}\right) = \left\{ -\frac{1}{8n}\frac{\sigma_{yyyy}}{\sigma_y{}^3\overline{Y}} + \frac{(n-3)}{8n(n-1)}\frac{\left(\sigma_{yy}\right)^2}{\sigma_y{}^3\overline{Y}} + \frac{1}{n}\left(\frac{\sigma_y}{\overline{Y}}\right)^3 + \frac{B_1{}^2\sigma_x{}^2\sigma_y}{n\overline{Y}^3} \right.$$
$$\left. -\frac{1}{2n}\frac{\sigma_{yyy}}{\sigma_y\overline{Y}^2} + \frac{B_1}{2n}\frac{\sigma_{xyy}}{\sigma_y\overline{Y}^2} - \frac{2B_1}{n}\frac{\sigma_{xy}\sigma_y}{\overline{Y}^3} \right\} + O\left(\frac{1}{n^2}\right), \tag{2.9}$$

and

$$MSE\left(\hat{\theta}_{y_2}\right) = \frac{B_1^2}{n}\frac{\sigma_x{}^2\sigma_y{}^2}{\overline{Y}^4} + \frac{1}{n}\left(\frac{\sigma_y}{\overline{Y}}\right)^4 + \frac{1}{4n}\frac{\sigma_{yyyy}}{\sigma_y{}^2\overline{Y}^2} - \frac{(n-3)}{4n(n-1)}\frac{\left(\sigma_{yy}\right)^2}{\sigma_y{}^2\overline{Y}^2} - \frac{1}{n}\frac{\sigma_{yyy}}{\overline{Y}^3}$$
$$\left. + \frac{B_1}{n}\frac{\sigma_{xyy}}{\overline{Y}^3} - \frac{2B_1}{n}\frac{\sigma_{xy}\sigma_y{}^2}{\overline{Y}^4} \right\} + O\left(\frac{1}{n^2}\right). \tag{2.10}$$

2.2.3 Estimator $\left(\hat{\theta}_{y_3}\right)$.

Standard textbooks in theory of sampling do not discuss the estimations of the population variance. Thus, regression estimators for the estimation of σ_y^2 (σ_{yy}) are not available in these textbooks. Following the discussion of the regression estimators for the population mean, we propose the regression estimator of $\sigma_{Y_R}{}^2$ as

$$\hat{\sigma}_{Y_R}{}^2 = s_y{}^2 + b_2\left(\sigma_x{}^2 - s_x{}^2\right). \tag{2.11}$$

where b_2 is the estimator of B_2, the regression coefficient of s_y^2 on s_x^2 which is given by

$$B_2 = \frac{E\left(s_{yy} - \sigma_{yy}\right)\left(s_{xx} - \sigma_{xx}\right)}{E\left(s_{xx} - \sigma_{xx}\right)^2} = \frac{\sigma_{xxyy} - \sigma_{xx}\sigma_{yy}}{\sigma_{xxxx} - \left(\sigma_{xx}\right)^2} + O\left(\frac{1}{n}\right). \tag{2.12}$$

The estimator b_2 is obtained by substituting the corresponding sample moments in (2.11).

Using this, the estimator $\hat{\theta}_{y_3}$ is proposed and is given by

$$\hat{\theta}_{y_3} = \frac{\left[s_y^{\,2} + b_2\left(\sigma_x^{\,2} - s_x^{\,2}\right)\right]^{\frac{1}{2}}}{\bar{y}}. \tag{2.13}$$

Further, bias and MSE for $\hat{\theta}_{y_3}$ are given as

$$bias\left(\hat{\theta}_{y_3}\right) = \left\{ -\frac{1}{8n}\frac{\sigma_{yyyy}}{\sigma_y^{\,3}\bar{Y}} + \frac{(n-3)}{8n(n-1)}\frac{\left(\sigma_{yy}\right)^2}{\sigma_y^{\,3}\bar{Y}} - \frac{B_2^{\,2}}{8n}\frac{\sigma_{xxxx}}{\sigma_y^{\,3}\bar{Y}} + \frac{B_2^{\,2}}{8n}\frac{(n-3)}{(n-1)}\frac{\left(\sigma_{xx}\right)^2}{\sigma_y^{\,3}\bar{Y}} \right. \tag{2.14}$$
$$\left. -\frac{1}{2n}\frac{\sigma_{yyy}}{\sigma_y\bar{Y}^2} + \frac{B_2}{2n}\frac{\sigma_{xxy}}{\sigma_y\bar{Y}^2} + \frac{B_2}{4n}\frac{\sigma_{xxyy}}{\sigma_y^{\,3}\bar{Y}} - \frac{B_2}{4n}\frac{\sigma_{xx}\sigma_{yy}}{\sigma_y^{\,3}\bar{Y}} \right\} + O\left(\frac{1}{n^2}\right),$$

and

$$MSE\left(\hat{\theta}_{y_3}\right) = \left\{ \frac{1}{n}\left(\frac{\sigma_y}{\bar{Y}}\right)^4 + \frac{1}{4n}\frac{\sigma_{yyyy}}{\bar{Y}^2\sigma_y^{\,2}} - \frac{(n-3)}{4n(n-1)}\frac{\left(\sigma_{yy}\right)^2}{\bar{Y}^2\sigma_y^{\,2}} + \frac{1}{4n}\frac{\sigma_{xxxx}B_2^{\,2}}{\sigma_y^{\,2}\bar{Y}^2} \right.$$
$$\left. -\frac{1}{4n}\frac{(n-3)}{(n-1)}\frac{\left(\sigma_{xx}\right)^2 B_2^{\,2}}{\sigma_y^{\,2}\bar{Y}^2} - \frac{1}{n}\frac{\sigma_{yyy}}{\bar{Y}^3} + \frac{1}{n}\frac{\sigma_{xxy}B_2}{\bar{Y}^3} - \frac{B_2}{2n}\frac{\sigma_{xxyy}}{\sigma_y^{\,2}\bar{Y}^2} + \frac{B_2}{2n}\frac{\sigma_{xx}\sigma_{yy}}{\sigma_y^{\,2}\bar{Y}^2} \right\} + O\left(\frac{1}{n^2}\right). \tag{2.15}$$

2.2.4 Estimator $\left(\hat{\theta}_{y_4}\right)$

This estimator is obtained by using regression estimators of σ_{yy} and $\bar{Y}$ respectively and is given by

$$\hat{\theta}_{y_4} = \frac{\left[s_y^{\,2} + b_2\left(\sigma_x^{\,2} - s_x^{\,2}\right)\right]^{\frac{1}{2}}}{\left[\bar{y} + b_1\left(\bar{X} - \bar{x}\right)\right]}. \tag{2.16}$$

The expressions for the bias and MSE of $\hat{\theta}_{y_4}$ are given by

$$bias\left(\hat{\theta}_{y_4}\right) = \left\{ \frac{1}{n}\left(\frac{\sigma_y}{\bar{Y}}\right)^3 + \frac{B_1^{\,2}\sigma_x^{\,2}\sigma_y}{n\bar{Y}^3} - \frac{B_2^{\,2}}{8n}\frac{\sigma_{xxxx}}{\sigma_y^{\,3}\bar{Y}} + \frac{B_2^{\,2}}{8n}\frac{(n-3)}{(n-1)}\frac{\left(\sigma_{xx}\right)^2}{\sigma_y^{\,3}\bar{Y}} - \frac{1}{8n}\frac{\sigma_{yyyy}}{\sigma_y^{\,3}\bar{Y}} + \frac{1}{8n}\frac{(n-3)}{(n-1)}\frac{\left(\sigma_{yy}\right)^2}{\sigma_y^{\,3}\bar{Y}} \right.$$
$$\left. -\frac{2B_1\sigma_{xy}\sigma_y}{n\bar{Y}^3} - \frac{1}{2n}\frac{\sigma_{yyy}}{\sigma_y\bar{Y}^2} + \frac{B_2}{2n}\frac{\sigma_{xxy}}{\sigma_y\bar{Y}^2} + \frac{B_1}{2n}\frac{\sigma_{xyy}}{\sigma_y\bar{Y}^2} - \frac{B_1B_2}{2n}\frac{\sigma_{xxx}}{\sigma_y\bar{Y}^2} + \frac{B_2}{4n}\frac{\sigma_{xxyy}}{\sigma_y^{\,3}\bar{Y}} - \frac{B_2}{4n}\frac{\sigma_{xx}\sigma_{yy}}{\sigma_y^{\,3}\bar{Y}} \right\} + O\left(\frac{1}{n^2}\right), \tag{2.17}$$

and

$$MSE\left(\hat{\theta}_{y_4}\right)=E\left(\hat{\theta}_{y_4}-\theta_{y_4}\right)=\left\{\frac{1}{n}\left(\frac{\sigma_y}{\bar{Y}}\right)^4+\frac{\sigma_x{}^2\sigma_y B_1{}^2}{n\bar{Y}^4}+\frac{1}{4n}\frac{\sigma_{yyyy}}{\sigma_y{}^2\bar{Y}^2}-\frac{(n-3)}{4n(n-1)}\frac{\left(\sigma_{yy}\right)^2}{\sigma_y{}^2\bar{Y}^2}+\frac{B_2{}^2}{4n}\frac{\sigma_{xxxx}}{\sigma_y{}^2\bar{Y}^2}-\frac{(n-3)}{4n(n-1)}\frac{\left(\sigma_{xx}\right)^2 B_2{}^2}{\sigma_y{}^2\bar{Y}^2}\right.$$

$$\left.-\frac{2B_1}{n}\frac{\sigma_{xy}\sigma_y{}^2}{\bar{Y}^4}-\frac{1}{n}\frac{\sigma_{yyy}}{\bar{Y}^3}+\frac{B_2}{n}\frac{\sigma_{xxy}}{\bar{Y}^3}+\frac{B_1}{n}\frac{\sigma_{xyy}}{\bar{Y}^3}-\frac{B_1B_2}{n}\frac{\sigma_{xxx}}{\bar{Y}^3}-\frac{B_2}{2n}\frac{\sigma_{xxyy}}{\sigma_y{}^2\bar{Y}^2}+\frac{B_2}{2n}\frac{\sigma_{xx}\sigma_{yy}}{\sigma_y{}^2\bar{Y}^2}\right\}+O\left(\frac{1}{n^2}\right).$$

(2.18)

2.2.5. Regression estimator $\left(\hat{\theta}_{y_5}\right)$

The preceding 3 estimators for population C.V consists of estimating the ratio $\frac{\sigma_y}{\bar{Y}}$ by using improved estimators for the numerator and denominator. We now propose regression estimator for this ratio. The basic logic is the same as in the case of estimation of the mean and some details are omitted. The regression estimator $\hat{\theta}_{y_5}$ is useful when we have the knowledge on the population mean $\bar{X}$ of the auxiliary variable. It is given by

$$\hat{\theta}_{y_5}=\frac{s_y}{\bar{y}}+b_3\left(\bar{X}-\bar{x}\right). \tag{2.19}$$

where b_3 is the estimate of B_3 which is given by

$$B_3=\frac{Cov\left(\frac{s_y}{\bar{y}},\bar{x}\right)}{V(\bar{x})}=\frac{\frac{\sigma_{xyy}}{2\bar{Y}\sigma_y}-\frac{\sigma_{xy}\sigma_y}{\bar{Y}^2}}{\sigma_{xx}}+O\left(\frac{1}{n}\right). \tag{2.20}$$

Further, the bias and MSE of $\hat{\theta}_{y_5}$ are given by

$$bias\left(\hat{\theta}_{y_5}\right)=\left\{\frac{1}{n}\left(\frac{\sigma_y}{\bar{Y}}\right)^3-\frac{1}{8n}\frac{\sigma_{yyyy}}{\sigma y^3\bar{Y}}+\frac{(n-3)}{8n(n-1)}\frac{\left(\sigma_{yy}\right)^2}{\sigma_y{}^3\bar{Y}}-\frac{1}{2n}\frac{\sigma_{yyy}}{\bar{Y}^2}\right\}+O\left(\frac{1}{n^2}\right), \tag{2.21}$$

and

$$MSE\left(\hat{\theta}_{y_5}\right)=\left\{\frac{1}{n}\sigma_x{}^2B_3{}^2+\frac{1}{n}\left(\frac{\sigma_y}{\bar{Y}}\right)^4+\frac{1}{4n}\frac{\sigma_{yyyy}}{\sigma_y{}^2\bar{Y}^2}-\frac{(n-3)}{4n(n-1)}\frac{\left(\sigma_{yy}\right)^2}{\sigma_y{}^2\bar{Y}^2}+\frac{2}{n}\frac{\sigma_{xy}B_3\sigma_y}{\bar{Y}^2}-\frac{1}{n}\frac{\sigma_{xyy}B_3}{\sigma_y\bar{Y}}-\frac{1}{n}\frac{\sigma_{yyy}}{\bar{Y}^3}\right\}+O\left(\frac{1}{n^2}\right).$$

(2.22)

2.2.6. Regression estimator ($\hat{\theta}_{y_6}$)

This estimator is useful when the information for the variance of the auxiliary variable is known and is given by,

$$\hat{\theta}_{y_6} = \frac{s_y}{\bar{y}} + b_4\left(\sigma_x{}^2 - s_x{}^2\right). \tag{2.23}$$

where b_4 is the estimate of B_4 which is given by

$$B_4 = \frac{Cov\left(\frac{s_y}{\bar{y}}, s_x{}^2\right)}{Var\left(s_x{}^2\right)} = \left\{\frac{\frac{\sigma_{xxyy}}{2\bar{Y}\sigma_y} - \frac{\sigma_{xx}\sigma_{yy}}{2\bar{Y}\sigma_y} - \frac{\sigma_{xxy}\sigma_y}{\bar{Y}^2}}{\sigma_{xxxx} - (\sigma_{xx})^2}\right\} + O\left(\frac{1}{n}\right). \tag{2.24}$$

Further, the bias and MSE for $\hat{\theta}_{y_6}$ are given by

$$bias\left(\hat{\theta}_{y_6}\right) = \left\{-\frac{1}{8n}\frac{\sigma_{yyyy}}{\bar{Y}\sigma_y{}^3} + \frac{(n-3)}{8n(n-1)}\frac{(\sigma_{yy})^2}{\bar{Y}\sigma_y{}^3} + \frac{1}{n}\left(\frac{\sigma_y}{\bar{Y}}\right)^3 - \frac{1}{2n}\frac{\sigma_{yyy}}{\bar{Y}^2\sigma_y}\right\} + O\left(\frac{1}{n^2}\right), \tag{2.25}$$

and

$$MSE\left(\hat{\theta}_{y_6}\right) = \left\{\frac{1}{4n}\frac{\sigma_{yyyy}}{\sigma_y{}^2\bar{Y}^2} - \frac{(n-3)}{4n(n-1)}\frac{(\sigma_{yy})^2}{\sigma_y{}^2\bar{Y}^2} + \frac{B_4{}^2}{n}\sigma_{xxxx} - \frac{(n-3)}{n(n-1)}B_4{}^2(\sigma_{xx})^2 + \frac{1}{n}\left(\frac{\sigma_y}{\bar{Y}}\right)^4 - \frac{B_4}{n}\left(\frac{\sigma_{xxyy}}{\sigma_y\bar{Y}}\right)\right.$$
$$\left. + \frac{B_4}{n}\frac{\sigma_{xx}\sigma_{yy}}{\sigma_y\bar{Y}} - \frac{1}{n}\frac{\sigma_{yyy}}{\bar{Y}^3} + \frac{2}{n}\frac{\sigma_{xxy}\sigma_y B_4}{\bar{Y}^2}\right\} + O\left(\frac{1}{n^2}\right). \tag{2.26}$$

2.2.7. Regression estimator $\left(\hat{\theta}_{y_7}\right)$

This regression estimator of the population C.V of the study variable uses the population C.V of the auxiliary variable. In many instances, although the information on population mean or variance of the auxiliary variable is not known, it is likely that the information on the population C.V of the auxiliary variable may be known. This is especially true with respect to sampling of forests, agricultural fields etc. The estimator is given by

$$\hat{\theta}_{y_7} = \frac{s_y}{\bar{y}} + b_5\left(\frac{\sigma_x}{\bar{X}} - \frac{s_x}{\bar{x}}\right), \tag{2.27}$$

where b_5 is the estimate of B_5 which is given by

$$B_5 = \frac{Cov\left(\frac{s_y}{\bar{y}}, \frac{s_x}{\bar{x}}\right)}{Var\left(\frac{s_x}{\bar{x}}\right)}$$

$$= \left\{ \frac{\frac{\sigma_{xxyy}}{4\bar{X}\bar{Y}\sigma_x\sigma_y} - \frac{\sigma_{xx}\sigma_{yy}}{4\bar{X}\bar{Y}\sigma_x\sigma_y} - \frac{\sigma_{xyy}\sigma_x}{2\sigma_y\bar{Y}\bar{X}^2} - \frac{1}{2}\frac{\sigma_{xxy}\sigma_y}{\sigma_x\bar{X}\bar{Y}^2} + \frac{\sigma_{xy}\sigma_x\sigma_y}{\bar{X}^2\bar{Y}^2}}{\frac{1}{4}\frac{\sigma_{xxxx}}{\sigma_x^2\bar{X}^2} - \frac{(\sigma_{xx})^2}{4\bar{X}^2\sigma_x^2} + \left(\frac{\sigma_x}{\bar{X}}\right)^4 - \frac{\sigma_{xxx}}{\bar{X}^3}} \right\} + O\left(\frac{1}{n}\right), \tag{2.28}$$

The expressions for the bias and MSE of $\hat{\theta}_{y_7}$ are given by

$$bias(\hat{\theta}_{y_7}) = \left\{ \frac{1}{n}\left(\frac{\sigma_y}{\bar{Y}}\right)^3 - \frac{1}{n}\left(\frac{\sigma_x}{\bar{X}}\right)^3 B_5 - \frac{1}{8n}\frac{\sigma_{yyyy}}{\sigma_y^3\bar{Y}} + \frac{1}{8n}\frac{(n-3)}{(n-1)}\frac{(\sigma_{yy})^2}{\sigma_y^3\bar{Y}} \right.$$
$$\left. + \frac{B_5}{8n}\frac{\sigma_{xxxx}}{\sigma_x^3\bar{X}} - \frac{(n-3)}{8n(n-1)}\frac{(\sigma_{xx})^2}{\sigma_x^3\bar{X}}B_5 - \frac{1}{2n}\frac{\sigma_{yyy}}{\sigma_y\bar{Y}^2} + \frac{1}{2n}\frac{\sigma_{xxx}}{\sigma_x\bar{X}^2}B_5 \right\} + O\left(\frac{1}{n^2}\right), \tag{2.29}$$

and

$$MSE(\hat{\theta}_{y_7}) = \left\{ \frac{1}{n}\left(\frac{\sigma_y}{\bar{Y}}\right)^4 + \frac{1}{n}\left(\frac{\sigma_x}{\bar{X}}\right)^4 B_5^2 + \frac{1}{4n}\frac{\sigma_{yyyy}}{\sigma_y^2\bar{Y}^2} - \frac{(n-3)}{4n(n-1)}\frac{(\sigma_{yy})^2}{\sigma_y^2\bar{Y}^2} + \frac{B_5^2\sigma_{xxxx}}{4n\sigma_x^2\bar{X}^2} - \frac{(n-3)}{4n(n-1)}\frac{B_5^2(\sigma_{xx})^2}{\sigma_x^2\bar{X}^2} \right.$$
$$\left. - \frac{2B_5\sigma_{xy}\sigma_x\sigma_y}{n\bar{X}^2\bar{Y}^2} - \frac{1}{n}\frac{\sigma_{yyy}}{\bar{Y}^3} + \frac{B_5}{n}\frac{\sigma_{xxy}\sigma_y}{\sigma_x\bar{X}\bar{Y}^2} + \frac{B_5}{n}\frac{\sigma_{xyy}\sigma_x}{\sigma_y\bar{X}^2\bar{Y}} - \frac{B_5^2\sigma_{xxx}}{n\bar{X}^3} - \frac{B_5}{2n}\frac{\sigma_{xxyy}}{\sigma_x\sigma_y\bar{X}\bar{Y}} + \frac{B_5}{2n}\frac{\sigma_{xx}\sigma_{yy}}{\sigma_x\sigma_y\bar{X}\bar{Y}} \right\} + O\left(\frac{1}{n^2}\right). \tag{2.30}$$

The Bias and MSE of these estimators are derived to the order of $O(n^{-1})$ by the authors using Taylor series expansion and higher order moments of sample mean and variance. These moments to the order of $O(n^{-1})$ are derived by the author and are given in Appendix A for the case of SRSWR and Appendix B for SRSWOR. The expression for the MSE of $\hat{\theta}_{y_1}$ coincides with the expressions derived by Kendall and Stuart (1977: p248) for the infinite population model. Thus, we notice that to the order of $O(1)$, $\hat{\theta}_{y_1}$ is unbiased.

2.3. SRSWOR

In the case of SRSWR, the sample variance ${s_y}^2$ is an estimate of the population variance ${\sigma_y}^2$. However, for SRSWOR design ${s_y}^2$ is an estimate of ${S_y}^2 = \frac{1}{N-1}\sum\left(Y_i - \bar{Y}\right)^2$.

Therefore, we define the population C.V for the study variable (y) as $\theta_y = \frac{S_y}{\bar{Y}}$.

The population C.V for the auxiliary variable *'x'* is similarly defined.

2.3.1. Usual Estimator $\left(\hat{\theta}_{y_1}\right)$

Following the same discussion for the case of SRSWOR, the usual estimator is given by

$$\hat{\theta}_{y_1} = \frac{s_y}{\bar{y}}, \tag{2.31}$$

where $\bar{y}$ is the sample mean and ${s_y}^2$ is the sample variance respectively. To simplify notations, no separate subscript is used in the case of SRSWOR and the context will make it clear whether the reference is WR or WOR schemes.
Further, the bias and MSE of $\hat{\theta}_{y_1}$ are given by

$$bias\left(\hat{\theta}_{y_1}\right) = \left\{-\frac{1}{8}\left(\frac{1}{n}-\frac{1}{N}\right)\frac{S_{yyyy}}{\bar{Y}{S_y}^3} + \frac{1}{8}\left(\frac{1}{n}-\frac{1}{N}\right)\frac{\left(S_{yy}\right)^2}{\bar{Y}{S_y}^3} + \left(\frac{1}{n}-\frac{1}{N}\right)\left(\frac{S_y}{\bar{Y}}\right)^3 - \frac{1}{2}\left(\frac{1}{n}-\frac{1}{N}\right)\frac{S_{yyy}}{S_y\bar{Y}^2}\right\} + O\left(\frac{1}{n^2}\right), \tag{2.32}$$

and

$$MSE\left(\hat{\theta}_{y_1}\right) = \left\{\frac{1}{4}\left(\frac{1}{n}-\frac{1}{N}\right)\frac{S_{yyyy}}{{S_y}^2\bar{Y}^2} - \frac{1}{4}\left(\frac{1}{n}-\frac{1}{N}\right)\frac{\left(S_{yy}\right)^2}{{S_y}^2\bar{Y}^2} + \left(\frac{1}{n}-\frac{1}{N}\right)\left(\frac{S_y}{\bar{Y}}\right)^4 - \left(\frac{1}{n}-\frac{1}{N}\right)\left(\frac{s_{yyy}}{\bar{Y}^3}\right)\right\} + O\left(\frac{1}{n^2}\right). \tag{2.33}$$

These expressions of bias and MSE are derived for the usual estimator in the case of SRSWOR by the authors and the expressions for MSE coincides with the expressions derived by Kendall and Stuart (1977) for the infinite population model, when $N \to \infty$.

2.3.2. Estimator $\left(\hat{\theta}_{y_2}\right)$

Using the regression estimator $\hat{\bar{Y}}_R = \bar{y} + b_1\left(\bar{X} - \bar{x}\right)$, the second estimator of θ_y is given by

$$\hat{\theta}_{y_2} = \frac{s_y}{\bar{y} + b_1\left(\bar{X} - \bar{x}\right)}. \tag{2.34}$$

The estimate of b_1 is B_1 which is given by $B_1 = \frac{S_{xy}}{S_{xx}}$. (2.35)

The bias and MSE of $\hat{\theta}_{y_2}$ is given by

$$bias\left(\hat{\theta}_{y_2}\right) = \left\{ -\frac{1}{8}\left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_{yyyy}}{S_y^{\,3}\bar{Y}} + \frac{1}{8}\left(\frac{1}{n} - \frac{1}{N}\right)\frac{\left(S_{yy}\right)^2}{S_y^{\,3}\bar{Y}} + \left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{S_y}{\bar{Y}}\right)^3 + \frac{B_1^{\,2}S_x^{\,2}S_y}{\bar{Y}^3}\left(\frac{1}{n} - \frac{1}{N}\right) \right.$$
$$\left. -\frac{1}{2}\left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_{yyy}}{S_y\bar{Y}^2} + \frac{B_1}{2}\left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_{xyy}}{S_y\bar{Y}^2} - \frac{2B_1}{\bar{Y}^3}\left(\frac{1}{n} - \frac{1}{N}\right)S_{xy}S_y \right\} + O\left(\frac{1}{n^2}\right), \tag{2.36}$$

and

$$MSE\left(\hat{\theta}_{y_2}\right) = \left\{ \frac{B_1^{\,2}S_x^{\,2}S_y^{\,2}}{\bar{Y}^4}\left(\frac{1}{n} - \frac{1}{N}\right) + \left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{S_y}{\bar{Y}}\right)^4 + \frac{1}{4}\left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_{yyyy}}{S_y^{\,2}\bar{Y}^2} - \frac{1}{4}\left(\frac{1}{n} - \frac{1}{N}\right)\frac{\left(S_{yy}\right)^2}{S_y^{\,2}\bar{Y}^2} \right.$$
$$\left. -\left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_{yyy}}{\bar{Y}^3} + B_1\left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_{xyy}}{\bar{Y}^3} - 2B_1\left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_{xy}S_y^{\,2}}{\bar{Y}^4} \right\} + O\left(\frac{1}{n^2}\right). \tag{2.37}$$

Similarly 5 other estimators of θ_y are also proposed in the case of SRSWOR scheme and the expressions for the Bias and MSE to the order of $O(n^{-1})$ are derived by the authors. To save space those expressions are not presented here and can be obtained by the authors.

3. First order comparison of the performance of the estimators.

From the expressions derived for the bias and MSE it is difficult to identify the estimators which have smaller bias and MSE. Thus, we have considered two data-sets to compare the estimators. They correspond to high value of population C.V and low value of population C.V respectively. The results corresponding to the data sets are described below.

3.1 Data Set (A)

This data set corresponds to payment on motor insurance in 63 geographical regions of Sweden (Source: Swedish Committee on Analysis of Risk Premium in Motor Insurance). The payment for all the claims in thousands of Swedish Kronas is taken as the study variable and the number of claims is taken as the auxiliary variable. Population characteristics for the variables x and y are reported below.

The population C.V's for the variables x and y are 1.01 and 0.88 respectively. The values of skewness (β_1) and kurtosis $(\gamma_2 = \beta_2 - 3)$ for the variable x are 2.32 and 6.85, while for the variable y the respective values are 1.63 and 3.33 respectively. Thus, the distributions for the x and y variables are highly right skewed and peaked. The population correlation coefficient between x and y is 0.91.

The bias and efficiency for the 7 estimators of the C.V (θ_y) for SRSWR and WOR designs for a sample of size n=20 are represented diagrammatically in Fig(3.1) and Fig(3.2). To save space the table is not reported here. It is clear that for the SRSWR scheme, the estimators having maximum efficiency are the regression estimator $\hat{\theta}_{y_7}$, where the information on the population C.V of the auxiliary variable is used (Efficiency (E($\hat{\theta}_{y_7}$)) =150.83), followed by the regression estimator $\hat{\theta}_{y_4}$ where the regression estimators of mean and variance are used to estimate θ_y with an efficiency of E($\hat{\theta}_{y_4}$)=150.50. The estimators having the least efficiency is $\hat{\theta}_{y_2}$ where improved estimator for the mean is used in the denominator of the ratio, $\hat{\theta}_y = \frac{\hat{\sigma}_y}{\hat{\bar{Y}}}$ (E($\hat{\theta}_{y_2}$)=18.41). The usual estimator $\hat{\theta}_{y_1}$ ranks fifth (E($\hat{\theta}_{y_1}$)=39.58), when the estimators are ranked in terms of efficiency in the descending order.

Figure (3.1) reflects the comparative performance of the seven estimators for SRSWR scheme represented through bar diagram. A similar conclusion is obtained for the SRSWOR scheme (see Figure (3.2)).

Figure 3.1: Efficiency of the seven estimators for the data set A under SRSWR

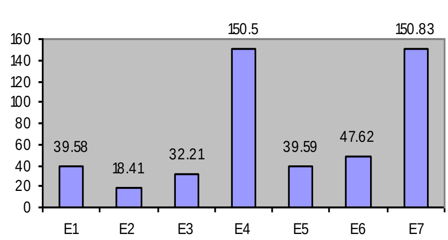
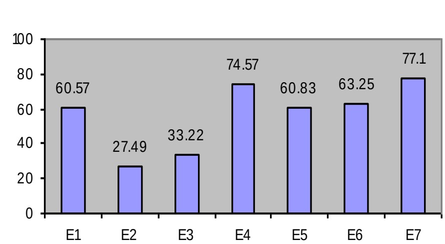


Figure 3.2: Efficiency of the seven estimators for the data set A under SRSWOR.



3.2. Data Set (B).

This real life data-set corresponds to cost of living index on grocery items and health care grouped by metropolitan areas of the United States (Source: Statistical Abstract of the United States, 120th edition). The cost of living index for health care is taken as the study variable and the cost living index for the grocery items is taken as the auxiliary variable. Population characteristics for the variables x and y are reported below.

The population C.V for the study variable 'y' is 0.10 and for the auxiliary variable 'x' is 0.074. The distributions of x and y variables are moderately right skewed. The distribution of variable x is mesokurtic $\left(\gamma_2 = \beta_2 - 3 = 0.047\right)$ and the distribution of variable y is slightly flat$\left(\gamma_2 = -0.30\right)$. The population correlation coefficient for the variables x and y is 0.84. The population size, N=45.

The bias and efficiency of the seven estimators of C.V *(θ_y)* are graphically represented using bar diagrams in Fig(3.3) and Fig(3.4) for the SRSWR and WOR designs. Since the population size is small the numerical calculation reported corresponds to a sample of size $n = 15$. It follows that for the SRSWR scheme, the maximum efficiency corresponds to the estimator $\hat{\theta}_{y_7}$ where the information on the C.V of the auxiliary variable is used (E($\hat{\theta}_{y_7}$)=6250.00), followed by the regression estimator $\hat{\theta}_{y_3}$ where improved estimator for the variance is used in the numerator of the ratio to estimate θ_y (E($\hat{\theta}_{y_3}$)=5571.03). The estimator having the least efficiency is $\hat{\theta}_{y_5}$ where information on the population mean $\overline{X}$ of the auxiliary variable are used (E($\hat{\theta}_{y_5}$)=3407.16), where as the usual estimator $\hat{\theta}_{y_1}$ ranks fourth, when the estimators are ranked in terms of efficiency in the descending order (E($\hat{\theta}_{y_1}$)=3695.49).

Figure (3.3) reflects the comparative performance of the estimators for SRSWR scheme. Proceeding in the same manner for SRSWOR, a similar conclusion is arrived (see Figure (3.4)).

Figure 3.3: Efficiency of the seven estimators for the data set B under SRSWR

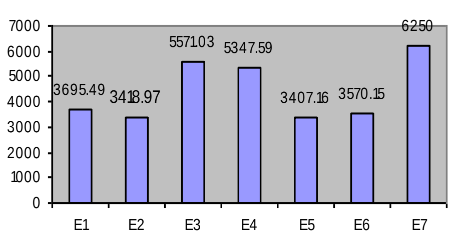
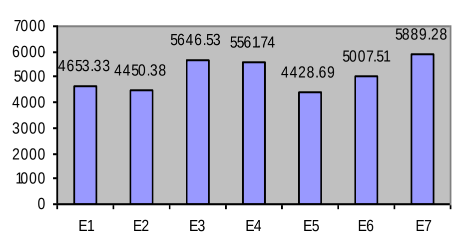


Figure 3.4: Efficiency of the seven estimators for the data set B under SRSW0R



In the preceding section we have compared the estimators via the asymptotic MSE. In the recent years the performance of the estimators are compared through the coverage probability of the confidence interval constructed from the estimator and the length of the confidence interval (Mahmoudvand and Hassani (2007)). Since the exact distributions of the estimators are difficult to tract analytically, we have carried out a simulation study to achieve the objective. Observations of size 'n' are generated from a bivariate normal distribution with parameters $\left(\mu_1, \sigma_1^2\right)$, $\left(\mu_2, \sigma_2^2\right)$ and ρ. For each sample the confidence interval is constructed using normal approximation to the distribution of the estimator. Each time the length of the confidence interval is also recorded using 10,000 simulations. The coverage probability and the average length of the confidence interval were recorded. Using the 10,000 simulated samples the MSE of the estimators are also computed. The comparison of the estimators through the average length of the confidence interval is valid only when they maintain the confidence level. Failure to maintain the confidence level only indicates that the normal approximation is not accurate to the sampling distribution of the estimators. In such cases the comparison is meaningful through the MSE.

The values of the C.V of the study variable used in the simulations were 0.1, 0.3, 0.5, 0.8, 1.0, and 2.0. For each fixed value of the study variable, a set of 4 values of C.V of the auxiliary variable are considered. They are 0.5, 1.0, 1.5 and 2 times the C.V of the study variable. The correlation co-efficient used in the simulation study are -0.9, -0.7, -0.5, -0.3, -0.1, 0, 0.1, 0.3, 0.5, 0.7, 0.9.

The sample sizes considered are n=100, 200. Only the confidence level used in the investigation=0.95.The total no. of configurations works out to be 6*4*11*2=528.

For each sample size and for a fixed value of C.V of the study variable the numerical values of the coverage probability are examined. In the present investigation a confidence interval is said to maintain the confidence level of 0.95, if the coverage probability exceeds 0.90 (approximately 5% error).

For each of the estimator at a fixed value of correlation co-efficient, the mean of the coverage probability is computed for the set of 4 values of C.V of the auxiliary variable, when in 3 or more cases the coverage probability exceeds 0.9. Whenever the confidence level is maintained the mean of the average length of the confidence interval is also obtained. When the estimators fail to maintain the confidence level attention is paid to the values of the MSE. After a careful scrutiny, two best estimators (satisfying the criterion of shortest length of the confidence interval/smaller MSE) are identified. They are $\hat{\theta}_4$ and $\hat{\theta}_7$. These are the best estimators for various configurations of the C.V of the auxiliary variable and the correlation co-efficient. For the other estimators, namely $\hat{\theta}_{y_2}$, $\hat{\theta}_{y_3}$, $\hat{\theta}_{y_5}$ (Regression estimator when information on mean of the auxiliary variable is used) and $\hat{\theta}_{y_6}$ (Regression estimator when information on variance of the auxiliary variable is used) no consistent pattern regarding the performance either in terms of coverage probability or MSE is emerged for the various configurations and to save space the results are not reported here.

Table 4.1 presents the average coverage probability of the confidence interval and average length of the confidence interval for the ratio type regression estimators ($\hat{\theta}_{y_4}$), the regression estimator ($\hat{\theta}_{y_7}$) where the information on the population C.V of the auxiliary variable is used and the sample C.V ($\hat{\theta}_{y_1}$).

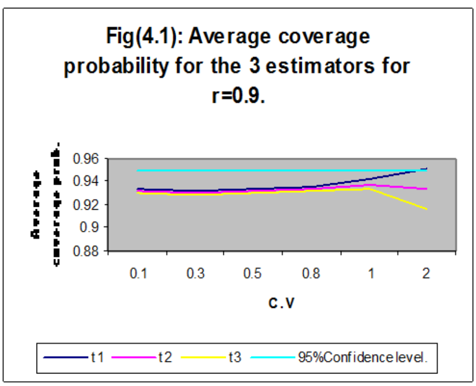
The regression estimators or the ratio type regression estimators do not maintain the coverage probability when the correlation co-efficient 'r' is low and thus the results are presented only when the correlation co-efficient is -0.9, -0.7, -0.5, 0.5, 0.7, 0.9. The result is presented in the table only when the sample size n=100 and the pattern of the results does not change for other sample size=200.

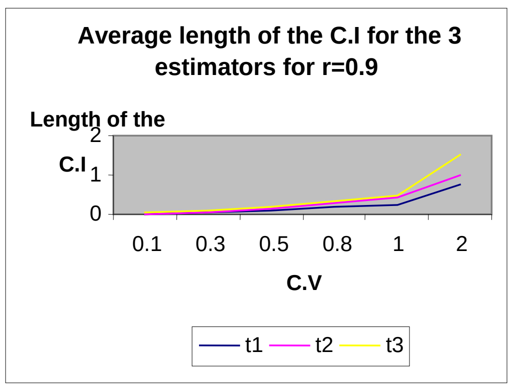
Table 4.1. Average coverage probability (in percentages) and average length of the C.V (in brackets) for α=0.05.

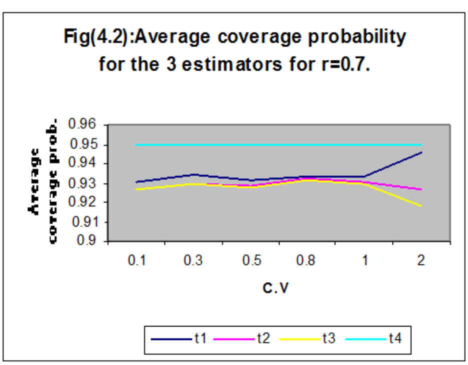
C.V of the study variable.		Correlation co-efficient (r) (n=100)					
		-0.9	-0.7	-0.5	0.5	0.7	0.9
	$\hat{\theta}_4$	0.9355(0.0159)	0.9315(0.0239)	0.9320(0.0268)	0.9326(0.0261)	0.9302(0.0240)	0.9335(0.0165)
	$\hat{\theta}_7$	0.9347(0.0182)	0.9257(0.0247)	0.9284(0.0271)	0.9277(0.0268)	0.9271(0.0241)	0.9324(0.0167)
0.1	$\hat{\theta}_1$	0.9342(0.0276)	0.9241(0.0281)	0.9273(0.0283)	0.9274(0.0287)	0.9269(0.0281)	0.9311(0.0271)
	$\hat{\theta}_4$	0.9334(0.0505)	0.9327(0.0737)	0.9313(0.0815)	0.9341(0.0819)	0.9343(0.0754)	0.9318(0.0513)
	$\hat{\theta}_7$	0.9313(0.0764)	0.9293(0.0839)	0.9271(0.0848)	0.9276(0.0821)	0.9298(0.0819)	0.9299(0.0631)
0.3	$\hat{\theta}_1$	0.9305(0.0895)	0.9278(0.0892)	0.9257(0.0883)	0.9261(0.0875)	0.9292(0.0883)	0.9287(0.0892)
	$\hat{\theta}_4$	0.9342(0.0897)	0.9344(0.1346)	0.9341(0.1569)	0.9338(0.1553)	0.9319(0.1376)	0.9337(0.0912)
	$\hat{\theta}_7$	0.9322(0.1592)	0.9314(0.1603)	0.9318(0.1638)	0.9290(0.1597)	0.9283(0.1469)	0.9323(0.1343)
0.5	$\hat{\theta}_1$	0.9305(0.1671)	0.9287(0.1678)	0.9307(0.1680)	0.9288(0.1686)	0.9278(0.1680)	0.9309(0.1673)
	$\hat{\theta}_4$	0.9349(0.1620)	0.9344(0.2516)	0.9330(0.2909)	0.9326(0.2929)	0.9334(0.2526)	0.9359(0.1676)
	$\hat{\theta}_7$	0.9333(0.3081)	0.9330(0.3099)	0.9326(0.3106)	0.9309(0.2989)	0.9323(0.2643)	0.9341(0.3026)
0.8	$\hat{\theta}_1$	0.9307(0.3253)	0.9294(0.3248)	0.9299(0.3256)	0.9287(0.3258)	0.9311(0.3253)	0.9323(0.3252)
	$\hat{\theta}_4$	0.9376(0.2224)	0.9379(0.3696)	0.9364(0.4280)	0.9330(0.4223)	0.9329(0.3643)	0.9418(0.2278)
	$\hat{\theta}_7$	0.9370(0.4458)	0.9315(0.4596)	0.9316(0.4636)	0.9304(0.4323)	0.9302(0.3965)	0.9376(0.4478)
1.0	$\hat{\theta}_1$	0.9329(0.4762)	0.9304(0.4795)	0.9294(0.4786)	0.9301(0.4780)	0.9292(0.4786)	0.9335(0.4769)
	$\hat{\theta}_4$	0.9594(0.7388)	0.9593(1.1144)	0.9498(1.3842)	0.9398(1.4031)	0.9458(1.2025)	0.9516(0.7778)
	$\hat{\theta}_7$	0.9343(0.9945)	0.9275(1.4393)	0.9219(1.4998)	0.9202(1.4893)	0.9268(1.4122)	0.9341(0.9987)
2.0	$\hat{\theta}_1$	0.9193(1.5441)	0.9200(1.5481)	0.9189(1.5463)	0.9176(1.5470)	0.9179(1.5471)	0.9161(1.5468)

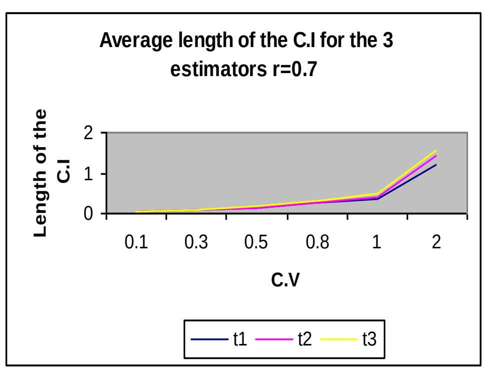
From the table it is clear that for the 2 estimators the coverage probability did not change either with the values of the correlation coefficient or with the values of C.V of the study variable. However, the length of the confidence interval for both the estimators steadily increases with the values of C.V of the study variable. When C.V=0.1, the average length of the confidence interval for the two estimators ($\hat{\theta}_{y_4}$ and $\hat{\theta}_{y_7}$) were respectively (0.0159, 0.0182), while for C.V=1.0, the respective values are (0.2224, 0.4458). The ratio of the length of the confidence interval to the value of C.V is approximately 16% when C.V=0.1, while it increased to 22% when C.V=1.0 for $\hat{\theta}_4$ and for $\hat{\theta}_7$ it is 18% for C.V=0.1 to 44% for C.V=1.0. For the sample C.V the average length of the confidence interval is larger compared to the other estimators. The ratio of the length of the confidence interval to the value of C.V for this estimator is 28% when CV=0.1, while it increased to 47% when C.V=1.0.

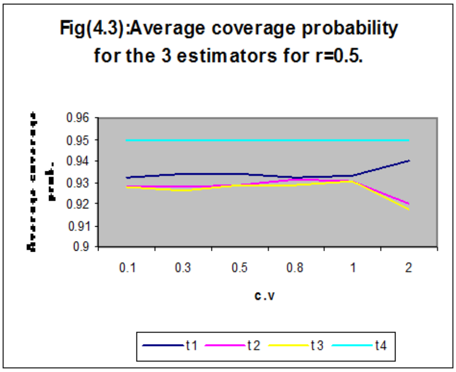
Fig (4.1), (4.2) and (4.3) represents the average length of confidence interval (right side) and average coverage probability (left side) versus C.V of the study variable for the 3 estimators.

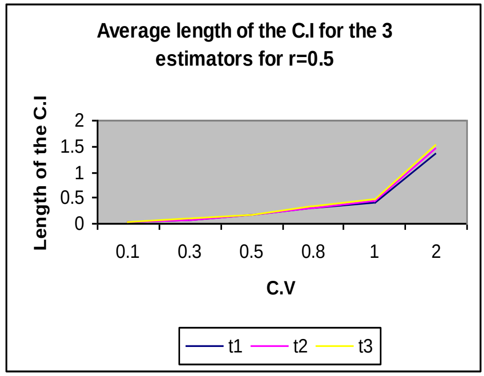












To obtain more accurate results we have compared the different estimators through their mean square error (MSE's). Table (4.2) presents the average mean square error of the ratio type regression estimator ($\hat{\theta}_{y_4}$), regression estimator ($\hat{\theta}_{y_7}$) where the information on the population C.V of the auxiliary variable is used and the sample C.V ($\hat{\theta}_{y_1}$) for low correlation co-efficient.

The MSE's of the ratio type regression estimator or the regression estimator increases steadily and it is observed that the estimators $\hat{\theta}_4$ and $\hat{\theta}_7$ has got the minimum MSE values when compared to $\hat{\theta}_1$ for the various configurations and when the correlation co-efficient is high. To save space the results are not reported when the correlation co-efficient is high. The results are presented only when the correlation co-efficient is -0.3, -0.1, 0, 0.1, 0.3 and also for the sample size n=100. The pattern of the results does not change for other sample size=200.

Table (4.2). Average MSE of the 3 estimators

C.V of the study variable.		Correlation co-efficient (r) (n=100)				
		-0.3	-0.1	0	0.1	0.3
0.1	$\hat{\theta}_4$	0.5126* 10^{-4}	0.5315* 10^{-4}	0.5219* 10^{-4}	0.4972* 10^{-4}	0.5185* 10^{-4}
	$\hat{\theta}_7$	0.5143* 10^{-4}	0.5362* 10^{-4}	0.5226* 10^{-4}	0.4978* 10^{-4}	0.5178* 10^{-4}
	$\hat{\theta}_1$	0.5109* 10^{-4}	0.5288* 10^{-4}	0.5161* 10^{-4}	0.4908* 10^{-4}	0.5133* 10^{-4}
0.3	$\hat{\theta}_4$	0.5616* 10^{-3}	0.5459* 10^{-3}	0.5487* 10^{-3}	0.5495* 10^{-3}	0.5249* 10^{-3}
	$\hat{\theta}_7$	0.5672* 10^{-3}	0.5463* 10^{-3}	0.5496* 10^{-3}	0.5510* 10^{-3}	0.5269* 10^{-3}
	$\hat{\theta}_1$	0.5596* 10^{-3}	0.5408* 10^{-3}	0.5423* 10^{-3}	0.5451* 10^{-3}	0.5209* 10^{-3}

Table (4.2). Average MSE of the 3 estimators (cont)

C.V of the study variable.		Correlation co-efficient (r) (n=100)				
		-0.3	-0.1	0	0.1	0.3
0.5	$\hat{\theta}_4$	0.0020	0.0018	0.0017	0.0019	0.0021
	$\hat{\theta}_7$	0.0021	0.0019	0.0018	0.0021	0.0024
	$\hat{\theta}_1$	0.0019	0.0017	0.0016	0.0018	0.0019
0.8	$\hat{\theta}_4$	0.0071	0.0078	0.0075	0.0073	0.0077
	$\hat{\theta}_7$	0.0072	0.0079	0.0076	0.0074	0.0078
	$\hat{\theta}_1$	0.0069	0.0074	0.0072	0.0071	0.0075
1.0	$\hat{\theta}_4$	0.0145	0.0153	0.0159	0.0164	0.0164
	$\hat{\theta}_7$	0.0154	0.0155	0.0161	0.0167	0.0169
	$\hat{\theta}_1$	0.0138	0.0148	0.0156	0.0156	0.0154
2.0	$\hat{\theta}_4$	0.2147	0.2582	0.3082	0.2432	0.2259
	$\hat{\theta}_7$	0.2286	0.2610	0.3085	0.2492	0.2289
	$\hat{\theta}_1$	0.1962	0.2430	0.3044	0.2357	0.2149

From the table which is reported here the conclusion that can be drawn is that for the 2 estimators the MSE of the estimator did not change either with the values of the correlation co-efficient or with the values of C.V of the study variable. However, the MSE for both the estimators steadily increases with the values of C.V of the study variable. When C.V=0.1, the average MSE for the two estimators ($\hat{\theta}_{y_4}$ and $\hat{\theta}_{y_7}$) were respectively (0.00005126, 0.00005143), while for C.V=1.0 the respective values of MSE are (0.0145, 0.0154). For the sample C.V, the respective values of average MSE are 0.00005109 for C.V=0.1 and 0.0138 for C.V=1.0. Thus, we notice that in the case of low correlation co-efficient, sample C.V performs better than the other 2 estimators.

5. Conclusions.

From the comparison of the asymptotic MSE's of the various estimators and the small sample comparison of the average length of the confidence interval, the conclusion that can be drawn is that the best estimators of C.V are the ratio type regression estimator, namely $\hat{\theta}_{y_4}$ and the regression estimator $\hat{\theta}_{y_7}$ where the information on population C.V of the auxiliary variable is used, when the auxiliary variable is correlated with the study variable. When there is low correlation the sample C.V emerges as the best estimator. In the estimation of population mean, regression estimator using the mean of the auxiliary variable emerges as the best estimator irrespective of the value of correlation co-efficient (see Murthy(1967)). However, in the estimation of population C.V, regression estimators performs well only when the correlation co-efficient is moderate or large. Among the two (regression / ratio type regression) estimators, the regression estimator using information on population C.V uses less information than the ratio type regression estimator. In many instances it is likely that information on population C.V of the auxiliary variable is available, while the individual values of population mean and variance are not available. Thus, we recommend the regression estimator for use when an auxiliary variable is properly chosen so as to be correlated with the study variable. In the absence of such auxiliary variable, it is safe to use sample C.V as the estimator.

Acknowledgements.

This paper is part of the first author's PhD work under the guidance of second author. For this paper, the author received the 'Jan Tinbergen Award' for the year 2007 and was presented during the 56^{th} session of ISI, in Lisbon, Portugal.

Appendix A

Population Moments in SRSWR

$$E\left(\bar{x}-\bar{X}\right)=0$$
$$E\left(\bar{y}-\bar{Y}\right)=0$$
$$E\left(\bar{x}-\bar{X}\right)^2=\frac{1}{n}\sigma_{xx}$$
$$E\left(\bar{y}-\bar{Y}\right)^2=\frac{1}{n}\sigma_{yy}$$
$$E\left(\bar{x}-\bar{X}\right)\left(\bar{y}-\bar{Y}\right)=\frac{1}{n}\sigma_{xy}$$
$$E\left(s_x{}^2-\sigma_x{}^2\right)=0$$
$$E\left(s_y{}^2-\sigma_y{}^2\right)=0$$
$$E\left(s_x{}^2-\sigma_x{}^2\right)^2=\frac{1}{n}\sigma_{xxxx}-\frac{(n-3)}{n(n-1)}\left(\sigma_{xx}\right)^2+O\left(\frac{1}{n}\right)$$
$$E\left(s_y{}^2-\sigma_y{}^2\right)^2=\frac{1}{n}\sigma_{yyyy}-\frac{(n-3)}{n(n-1)}\left(\sigma_{yy}\right)^2+O\left(\frac{1}{n}\right)$$

$$E\left(s_x^{\;2}-\sigma_x^{\;2}\right)\left(\bar{x}-\bar{X}\right)=\frac{1}{n}\sigma_{xxx}+O\left(\frac{1}{n}\right)$$

$$E\left(s_x^{\;2}-\sigma_x^{\;2}\right)\left(\bar{y}-\bar{Y}\right)=\frac{1}{n}\sigma_{xxy}+O\left(\frac{1}{n}\right)$$

$$E\left(s_y^{\;2}-\sigma_y^{\;2}\right)\left(\bar{x}-\bar{X}\right)=\frac{1}{n}\sigma_{xyy}+O\left(\frac{1}{n}\right)$$

$$E\left(s_x^{\;2}-\sigma_x^{\;2}\right)\left(\bar{x}-\bar{X}\right)=\frac{1}{n}\sigma_{xxx}+O\left(\frac{1}{n}\right)$$

$$E\left(s_x^{\;2}-\sigma_x^{\;2}\right)\left(\bar{y}-\bar{Y}\right)=\frac{1}{n}\sigma_{xxy}+O\left(\frac{1}{n}\right)$$

$$E\left(s_y^{\;2}-\sigma_y^{\;2}\right)\left(\bar{x}-\bar{X}\right)=\frac{1}{n}\sigma_{xyy}+O\left(\frac{1}{n}\right)$$

$$E\left(s_y^{\;2}-\sigma_y^{\;2}\right)\left(\bar{y}-\bar{Y}\right)=\frac{1}{n}\sigma_{yyy}+O\left(\frac{1}{n}\right)$$

$$E\left(s_x^{\;2}-\sigma_x^{\;2}\right)\left(s_y^{\;2}-\sigma_y^{\;2}\right)=\frac{1}{n}\sigma_{xxyy}-\frac{1}{n}\sigma_{xx}\sigma_{yy}+O\left(\frac{1}{n}\right)$$

Note: $\sigma_{yy}=\dfrac{1}{N}\sum_{i}^{N}\left(Y_i-\bar{Y}\right)^2 \qquad \sigma_{xy}=\dfrac{1}{N}\sum_{i=1}^{N}\left(X_i-\bar{X}\right)\left(y_i-\bar{Y}\right)$

$$\sigma_{yyyy}=\frac{1}{N}\sum_{i=1}^{N}\left(Y_i-\bar{Y}\right)^4 \qquad \sigma_{\mathrm{yyy}}=\frac{1}{N}\sum_{i=1}^{N}\left(Y_i-\bar{Y}\right)^3$$

$$\sigma_{xxy}=\frac{1}{N}\sum_{i=1}^{N}\left(X_i-\bar{X}\right)^2\left(Y_i-\bar{Y}\right)$$

In the similar manner the other moments are defined. To save space the expressions are not given.

Appendix B

For the case of SRSWOR, the expressions for the moments are defined in a similar manner as in Appendix A. But Population variance σ_y^2 (σ_{yy}) is replaced by S_y^2 (S_{yy}). To save space the expressions are not presented here.

REFERENCES:

AHMED, S.E (2002): Simultaneous estimation of Co-efficient of Variation. *Journal of Statistical Planning and Inference*. 104, *31-51*.

ANANTHAKRISHNAN, R. and SOMAN, M.K. (1989): Statistical distribution of daily rainfall and its association with the co-efficient of variation of rainfall series. *International Journal of Climatology* 9, *485-500*.

BREUNIG, R.(2001): An almost unbiased estimator of the co-efficient of variation. *Economics Letters*, 70(1), *15-19*.

BRUS, D.J and RIELE W.J. Te (2001): Design-based regression estimators for spatial means of soil properties, *Geoderma*, 104, *257-279*.

CHATURVEDI, A. and RANI, U(1996): Fixed width confidence interval estimation of Inverse co-efficient of variation in normal population. *Microelectronics and Reliabiliy*, 36, *1305-1308*.

DAS, A.K. and TRIPATHI, T.P. (1981 a): Sampling strategies for coefficient of variation using knowledge of the mean using an auxiliary character. *Tech. Rep. Stat. Math. 5/81*. ISI, Calcutta.

DAS, A.K. and TRIPATHI, T.P. (1981 b): A class of estimators for co-efficient of variation using knowledge on coefficient of variation of an auxiliary character. Paper presented at annual conference of *Ind. Soc. Agricultural Statistics. Held at New Delhi, India.*

DE, P., GHOSH, J.B., and WELLS, C.E.(1996): Scheduling to minimize the co-efficient of variation. *International Journal of Production Economics*, 44, *249-253*.

FIELLER, E.C. (1932): A numerical test of the adequacy of A. T. McKay's approximation. *J. Roy. Stat. Soc.* 95, pp. *699–702*.

HENDRICKS, W.A. and ROBEY, W.K.(1936): The sampling distribution of the Co-efficient of variation. *Annals. Math. Stat.* (7), *129-132*.

KENDALL, M. and STUART, A. (1977): *The advanced Theory of Statistics*, 1, 4th edition. London: Charles Griffin and Co.

Kordonsky, Kh.B and Gertsbakh. I (1997): Multiple Time Scales and Life time Coefficient of Variation: Engineering Applications. *Lifetime Data Analysis*. 3, *139-156*.

MAHMOUDVAND, R. and HASSANI, H. (2007): Two new confidence intervals for the C.V in a normal distribution. *Journal of Applied Statistics,* 36: 4,*429-442*.

MCKAY, A.T. (1932): Distribution of the Co-efficient of variation and the extended 't' distribution. J.Royal.Stat.Soc.95, *696-698*.

MUKHOPADHYAY, PARIMAL (2008): *Theory and Methods of Survey Sampling*, Second edition, Prentice Hall of India, New Delhi.

MURTHY M.N. (1967): *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta.

NAIRY, S.K. and RAO, A.K. (2003): Tests for Coefficients of variation of normal population. *Communications in Statistics-Simulation and Computation*, 32(3), *641-661*.

NG, C.K.(2006): Performance of three methods of interval estimation of the co-efficient of variation. *InterStat*.

PANICHKITKOSOLKUL, W. (2009): Improved confidence intervals for a Co-efficient of variation of a normal distribution. *Thailand statistician*, 7(2), *193-199.*

PATEL, P. A. and SHAH RINA. (2009): A Monte Carlo comparison of some suggested estimators of Co-efficient of variation in finite population. *Journal of Statistics sciences, 1(2), 137-147*.

PRADHAN, B.K. (2010): Regression type estimators of finite population variance under multiphase sampling, *Bull. Malays. Math. Sci. Soc*, 33(2), *345-353.*

PEARSON, E.S.(1932): Comparison of A.T. McKay's approximation with experimental sampling results. J.Royal. Stat. Soc. 95, *703-704*.

RAJYAGURU, A. and GUPTA, P.C.(2002): On the estimation of Co-efficient of variation for finite population-I, *Journal of Statistical research, Dhaka*.

RAO., A.K, and Bhatta, A.R.S. (1989): A note on test for Co-efficient of Variation. *Calcutta Statistical Association Bulletin*, 38, *151-152*.

SINGH,H.P, SINGH,S and KIM, J.M.(2010): Efficient use of auxiliary variables in estimating finite population variance in two-phase sampling. *Communications of the Korean Statistical Society*, 17(2), *165-181*.

SAHOO, L.N; SAHOO, R.K and SENAPATHI, S.C (2003): Predictive estimation of Finite Population mean using Regression type estimators in double sampling procedures. *Journal of Statistical Research*, 37, *291-296.*

SINGH, M. (1993): Behavior of sample Co-efficient of Variation drawn from several distributions. *Sankhya*. *B*. 55, *65 – 76*.

SRIVASTAVA, S.K.(1980): A class of estimators using auxiliary information in sample surveys. *Canad. J. Statist*, 8(2), *253-254.*

TRIPATHI, T.P. SINGH, H.P. and UPADHYAYA, L.N. (2002): A general method of estimation and its application to the estimation of co-efficient of variation. *Statistics in Transition*, 5(6), *887-909.*

VERMA, M.R. (2008): Approximately optimum stratification for ratio and regression methods of estimation. *Applied Mathematics Letters*, 21(2),*200-207*.

VERRIL, S. (2003): Confidence bounds for normal and log-normal distribution co-efficient of variation, *Research paper,EPL-RP-609,Madison,Wisconsin,US*