Chapter 5

Estimation of the Coefficient of Variation - I

5.1 Introduction

The coefficient of variation (CV), which is the ratio of the standard deviation to the mean, is a relative measure of dispersion found to be very useful in many situations. Extensive work on estimation of CV for infinite populations has been made, see, e.g., Miller and Karson (1977), Hamer et al. (1995), Ahn (1995), Gong & Li (1999), Schwartz et al. (2000), Aston et al. (2003), among others.

Estimation of CV in finite population is also very important; it is often used as a measure of precision of an estimator. The real use of an estimator of a finite population CV would be in determining a sample size for a future study. This is evident from the following example

**Example:** Suppose that one wishes to estimate a population parameter $\theta$ - for example, the population mean or total-with an estimator $\hat{\theta}$. One would wish the estimate to be close to the true value with high probability. Specifying a maximum allowable difference $d$ between the estimate and the true value and allowing for a small probability $\alpha$ that the error may exceed that difference, the object is to choose a sample size $n$ such that

$$P(|\hat{\theta} - \theta| > d) < \alpha$$
If the estimator $\hat{\theta}$ is an unbiased normally distributed estimator of $\theta$ then
\[
\frac{\hat{\theta} - \theta}{\sqrt{\text{var}(\hat{\theta})}}
\]
has a standard normal distribution. Letting $z$ denote the upper $\alpha/2$ point of the standard normal distribution yields
\[
P\left( \frac{\hat{\theta} - \theta}{\sqrt{\text{var}(\hat{\theta})}} > Z \right) = P\left( |\hat{\theta} - \theta| > z \sqrt{\text{var}(\hat{\theta})} \right) = \alpha
\]
The variance of the estimator $\hat{\theta}$ decreases with increasing sample size $n$, so that the inequality above will be satisfied if we can choose $n$ large enough to make $z \sqrt{\text{var}(\hat{\theta})} \leq d$.

With simple random sampling, the sample mean $\bar{y}$ is an unbiased estimator of the population mean $\mu$ with variance $\text{var}(\bar{y}) = \frac{(N-n)\sigma^2}{Nn}$. Setting
\[
z \sqrt{\left( \frac{N-n}{N} \right) \frac{\sigma^2}{n}} = d
\]
and solving for $n$ gives the necessary sample size:
\[
n = \frac{1}{d^2 / z^2 \sigma^2 + 1/N} = \frac{1}{1/n_0 + 1/N}
\]
where $n_0 = \frac{z^2 \sigma^2}{d^2}$.

If the population size $N$ is large relative to the sample size $n$, so that the finite population correction factor can be ignored, the formula for sample size simplifies to $n_0$. 

99
If instead of controlling the absolute error \(d\), one is concerned with the relative error \(r\)-that is, the difference between the estimate and the true value, divided by the true value- the criterion to be met is

\[
P \left( \left| \frac{\hat{\theta} - \theta}{\theta} \right| > r \right) < \alpha
\]

Or, equivalently,

\[
P(\hat{\theta} - \theta > r\theta) < \alpha
\]

where \(\theta\) represents, either the population mean or the population total.

To estimate the population mean \(\mu\) to within \(r\mu\) of the true value or to estimate the population total \(\tau\) to within \(r\tau\) of the true value with probability \(1 - \alpha\), the sample size formula is

\[
n = \frac{1}{r^2 \mu^2 / z^2 \sigma^2 + 1/N}
\]

Letting \(\gamma\) denote the coefficient of variation for the population (i.e., \(= \sigma / \mu\)) the sample size formula may be written

\[
n = \frac{1}{r^2 / z^2 \gamma^2 + 1/N}
\]

Thus, the coefficient of variation is the population quantity on which sample size depends when the desire is to control relative precision.

This chapter presents several alternative estimators of a finite population coefficient of variation when samples are selected using SRSWOR and when complete auxiliary information is available. To compare these estimators empirically a small scale simulation study is carried out using real populations. The problem of estimation of CV has been discussed by Das & Tripathi (1981a,b), Rajayaguru and Gupta (2002, 2006), Tripathi et al. (2002), among other.
We seek to estimate the population coefficient of variation defined by

$$C_y = \left( \frac{S_y}{\bar{Y}} \right) \times 100$$  \hspace{1cm} (5.1)

on the basis of a random sample drawn from U according to SRSWOR, where $S_y$ is the positive square root of

$$S_y^2 = \sum_i \sum_{j \in U} C_{ij} y_i y_j$$

with $C_{ij} = 1/N$ for $i = j$ and $C_{ij} = -1/N(N-1)$ for $i \neq j$.

The customary estimator of $C_y$ under SRSWOR is the sample CV defined by

$$\hat{C}_y = \left( \frac{s_y}{\bar{y}} \right) \times 100$$  \hspace{1cm} (5.2)

where $s_y$ is the positive-square root of

$$s_y^2 = \sum_i \sum_{j \in S} c_{ij} y_i y_j$$

with $c_{ij} = 1/n$ for $i = j$ and $c_{ij} = -1/n(n-1)$ for $i \neq j$.

The approximate bias and variance of $C_y$ are given by (see 1.9 on p-32)

$$Bias(\hat{C}_y) = \frac{\mu_{20}}{n\mu_{01}} \left[ \frac{\mu_{02}}{\mu_{01}} - \frac{1}{8} \left( \frac{\mu_{04}}{\mu_{02}} \right) - \frac{\mu_{03}}{2\mu_{01}\mu_{02}} \right]$$

$$= \left( \frac{1-f}{n} \right) C_y^2 \left( C_y^2 + \frac{1}{8} \lambda_{40} - \frac{1}{2} C_y \lambda_{30} \right)$$  \hspace{1cm} (5.3)

and

$$V(\hat{C}_y) = \frac{\mu_{20}}{n\mu_{01}^2} \left[ \frac{\mu_{02}}{\mu_{01}} + \frac{1}{4} \left( \frac{\mu_{04}}{\mu_{02}} - 1 \right) - \frac{\mu_{03}}{\mu_{01}\mu_{02}} \right]$$

$$= \left( \frac{1-f}{n} \right) C_y^2 \left( C_y^2 + \frac{1}{4} \lambda_{40} - C_y \lambda_{30} \right)$$  \hspace{1cm} (5.4)
where $\mu_1 = \bar{Y}, \mu_0 = \bar{X}, \mu_{rs} = \frac{1}{n} \sum_s (y_i - \bar{y})^r (x_i - \bar{x})^s, r, s = 1, 2, \ldots, \left( \lambda_{rs} = \frac{\mu_{rs}}{\mu_{rs}^2} \right), r, s = 0, 1, 2, \ldots,$

Das and Tripathi (1981 a, b) presented two classes of estimators for $C_y$ in different situations which are

\[
\hat{C}_{yD1} = \frac{\hat{C}_y + \lambda_1 (\bar{X} - \bar{x})}{[\bar{x} + \lambda_2 (\bar{X} - \bar{x})]^\alpha_1} (\bar{X})^{\alpha_1}, \text{ when } \bar{X} \text{ is known} \tag{5.5}
\]

and

\[
\hat{C}_{yD2} = \frac{\hat{C}_y + \delta_1 (C_x^2 - \hat{C}_x^2)}{[\hat{C}_x^2 + \delta_2 (C_x^2 - \hat{C}_x^2)]^{\alpha_2}} (C_x^2)^{\alpha_2}, \text{ when } C_x \text{ is known} \tag{5.6}
\]

where $\lambda$’s and $\alpha$’s are suitably chosen constants. These classes of estimators include many estimators, viz.,

- Ratio-estimator: $\hat{C}_{yR} = \hat{C}_y \frac{C_x}{\hat{C}_x}$ \( \tag{5.7} \)

- Difference estimator: $\hat{C}_{yD} = \hat{C}_y + k (C_x - \hat{C}_x)$ \( \text{ (k is known)}, \tag{5.8} \)

- Regression estimator: $\hat{C}_{yREG} = \hat{C}_y + b (C_x - \hat{C}_x)$ \( \text{ (b is a sample fraction)}. \tag{5.9} \)

Using Result 1.2 (given in chapter 1, p-27) we can obtain the optimum approximate variances of $\hat{C}_{yD1}$ and $\hat{C}_{yD2}$ as

\[
V(\hat{C}_{yD1}) \approx V(\hat{C}_y) - \frac{1-f}{n} C_y^2 \left[ \frac{1}{2} \lambda_{21} - C_y \lambda_{11} \right]^2 \tag{5.10}
\]

and

\[
V(\hat{C}_{yD2}) \approx V(\hat{C}_y) - \frac{1-f}{n} C_y^2 \left[ \frac{1}{2} \lambda_{21} C_x - \frac{1}{2} C_x \lambda_{12} - \frac{1}{2} \lambda_{11} \lambda_{12} - \frac{1}{4} \lambda_{22} \right]^2 \left[ C_x^2 + \frac{1}{4} \lambda_{04} - C_x \lambda_{03} \right] \tag{5.11}
\]
Tripathi et al. (2002) suggested the class of difference-type estimators of $C_y$ as

$$\hat{C}_{yD3} = \hat{C}_y + \delta_1(\bar{x} - \bar{y}) + \delta_2(s_x^2 - s_x^2)$$

where $\delta_1$ and $\delta_2$ are suitably chosen constants. Further, they compared various estimators by taking only two real populations and found that the estimators $\hat{C}_{yD2}$ (with optimum $\beta_2$) and $\hat{C}_{yD3}$ performed well. The latter was slightly more efficient than the former.

### 5.2 A system of estimators

Regression estimation is a powerful technique for estimating finite population means of survey variables when the population means of a set of auxiliary variables is known. Two well-known type of regression estimators have been available, namely the generalized regression estimator (GREG) (Särndal et. al., 1992) and the optimal estimator (OPE) (Montenari and Ranalli, 2002). Shah and Patel (1996), using regression methods of estimation, have suggested a system of estimators for a finite population variance which includes many potentially interesting estimators. Motivated by this, we define a system of estimators for estimating $C_y$ by creating the regression estimators for the numerator and the denominator of $C_y$ and then using the ratio of the two as an estimator of $C_y$ as

$$\hat{C}_{ys} = \frac{s_y^2 + \hat{\gamma}(s_x^2 - s_x^2)^{1/2}}{\bar{y} + \hat{\beta}(\bar{X} - \bar{x})} \times 100\% \quad (5.12)$$

As the special cases of this system, some interesting estimators are discussed below.

#### 5.2.1 The Optimal-type estimator

The difference estimators of $\bar{Y}$ and $S_y^2$ under SRSWOR are given by

$$\hat{Y}_D = \bar{y} + \beta(\bar{X} - \bar{x})$$
where $\beta$ and $\gamma$ are known constants. The optimal values of $\beta$ and $\gamma$ can be obtained by minimizing the variances of $\hat{Y}_D$ and $\hat{S}_{YD}^2$ respectively.

Consequently, the optimal-type estimator of $C_y$ is defined by replacing $S_y^2$ and $\bar{y}$ by their design-based optimal estimators which is given by

$$\hat{C}_{yo} = \frac{s_y^2 + \hat{\gamma}_0 (S_x^2 - s_x^2)^{1/2}}{\bar{y} + b (\bar{X} - \bar{x})} \times 100$$

(5.13)

where

$$b = \frac{s_{xy}}{s_x^2} \quad \text{and} \quad \hat{\gamma}_0 = \frac{\text{Cov}(S_x^2, s_y^2)}{\bar{y}(s_x^2)} = \frac{k_1 m_{22} + k_2 [2m_{11}^2 + m_{20}m_{02}] - k_3 m_{20}m_{02}}{k_1 m_{04} + (3k_1 - k_3)m_{02}^2}$$

with

$$k_1 = \frac{N(N-n)(Nn-n-N-1)}{n(n-1)(N-1)(N-2)(N-3)}, \quad k_2 = \frac{N(N-n)(Nn-n-1)}{n(n-1)(N-1)(N-2)(N-3)}$$

$$k_3 = \frac{N(N-n)}{(n-1)(N-1)^2(N-2)}$$

$$\mu_{rs} = \frac{\sum_s (y_i - \bar{y})^r (x_i - \bar{x})^s}{n}, r, s = 0, 1, 2, 3, 4$$

5.2.2 The Calibration estimator

Calibration has become a widely used procedure for estimation in sample surveys. It uses auxiliary information to produce efficient estimators. An efficient calibration estimator can be obtained if $x_i$ explains a significant part of the variability of $y_i$.

Deville and Särndal (1992) introduced calibration approach to estimate the population total. The basic idea was to minimize a measure of distance between a
given set of initial weights and the required calibration weights subject to the calibration constraint to obtain set of calibration weights (see, for details, § 1.3.4). Singh et al. (1999), Patel and Chaudhari (2003) extended this approach to obtain the estimator of $S_y^2$. Motivated by this the calibration type estimator of $C_y$, is given by

$$
\hat{C}_{yc} = \frac{\sqrt{s_y^2 + \hat{y}_c(S_x^2 - s_x^2)}}{\bar{y} + b(\bar{X} - \bar{x})} \times 100
$$

(5.14)

where $b$ is the least square estimator of the slope coefficient in the regression of $y_i$ and $x_i$ given above and

$$
\hat{y}_c = \frac{c_1 \sum_i x_i^2 y_i^2 + c_2 \sum_i (x_i y_i)(x_j y_j)}{c_1 \sum_i x_i^4 + c_2 \sum_i (x_i^2 x_j^2)}
$$

5.2.3 The Isaki-type estimator

Isaki (1983) suggested the estimator of $S_y^2$ as $\hat{s}_y^2 + \hat{y}_l(S_x^2 - s_x^2)$ where $\hat{y}_l = \left(\frac{s_{xy}}{s_x^2}\right)^2$. The main idea is to adjust $S_y^2$ in such a manner as to reduce its bias through the use of auxiliary variable. The Isaki-type estimator of $C_y$ is then readily defined as

$$
\hat{C}_{yl} = \frac{\sqrt{s_y^2 + \hat{y}_l(S_x^2 - s_x^2)}}{\bar{y} + b(\bar{X} - \bar{x})} \times 100
$$

(5.15)

5.3 Approximate variance of the system of estimators

In order to obtain the minimum attainable approximate variance of $\hat{C}_{ys}$, we denote
\[ \hat{C}_{ys} = g(\hat{\bar{y}}, \bar{x}, s_{\hat{y}}^2, s_{\hat{x}}^2, b, \hat{\gamma}) = g(\hat{\theta}) \]  
(5.16)

\[ C_y = g(\bar{y}, \bar{x}, s_y^2, S_x^2, \beta, \gamma) = g(\theta) \]  
(5.17)

where \( \hat{\theta}_1 = \bar{y}, \hat{\theta}_2 = \bar{x}, \hat{\theta}_3 = s_y^2, \hat{\theta}_4 = s_x^2, \hat{\theta}_5 = b, \hat{\theta}_6 = \hat{\gamma} \) and the corresponding population parameters are denoted by \( \theta_i, i = 1, \ldots, 6 \), respectively.

To derive the approximate bias and variance of \( g(\hat{\theta}) \) we need to consider the following expected values which are valid up to terms of order \( n^{-1} \):

Using Appendix II

\[ E[\hat{\theta} = \theta] \]

and

\[
\frac{1-f}{n} \sum (\hat{\theta} - \theta) (\hat{\theta} - \theta) \approx \begin{pmatrix}
\mu_{20} & \mu_{11} & \mu_{30} & \mu_{12} & * & *
\mu_{02} & \mu_{21} & \mu_{03} & * & *
\mu_{40} & \mu_{22} & * & * & *
\mu_{04} & * & * & * & *
* & * & * & * & *
\end{pmatrix}
\]  
(5.18)

where “*” indicates the quantity need not to know.

From (5.16), the following are easily verified

\[
\left( \frac{\partial C_{ys}}{\partial \bar{y}} \right)^2_{\hat{\theta} = \theta} V(\bar{y}) = \left( -\frac{(S_y^2)^{1/2}}{\bar{y}^2} \right)^2 V(\bar{y}) = \frac{S_y^2}{\bar{y}^4} V(\bar{y})
\]

\[
\left( \frac{\partial C_{ys}}{\partial \bar{x}} \right)^2_{\hat{\theta} = \theta} V(\bar{x}) = \left( -\frac{(S_y^2)^{1/2}}{\bar{y}^2} \beta \right)^2 V(\bar{x}) = \beta^2 \frac{S_y^2}{\bar{y}^4} V(\bar{x})
\]

\[
\left( \frac{\partial C_{ys}}{\partial s_y^2} \right)^2_{\hat{\theta} = \theta} V(s_y^2) = \frac{1}{4 \bar{y}^2 S_y^2} V(s_y^2)
\]
\[
\left( \frac{\partial C_{ys}}{\partial s_{\bar{y}}^2} \right)_{\theta=\bar{y}} V(s_{\bar{y}}^2) = \frac{\gamma^2}{4Y^2S_{\bar{y}}^2} V(s_{\bar{y}}^2)
\]

\[
\left( \frac{\partial C_{ys}}{\partial \beta} \right)_{\theta=\bar{y}} V(\beta) = 0
\]

\[
\left( \frac{\partial C_{ys}}{\partial \gamma} \right)_{\theta=\bar{y}} V(\gamma) = 0
\]

\[
\left( \frac{\partial C_{ys}}{\partial \bar{y}} \right) \frac{\partial C_{ys}}{\partial \bar{x}} \left( \frac{\partial C_{ys}}{\partial \bar{y}} \right)_{\theta=\bar{y}} \text{cov}(\bar{y}, \bar{x}) = -\frac{(S_{\bar{y}}^2)^{1/2}}{\bar{Y}^2} \frac{(S_{\bar{y}}^2)^{1/2}}{\bar{Y}^2} \beta \text{cov}(\bar{y}, \bar{x}) = -\beta \frac{S_{\bar{y}}^2}{\bar{Y}^4} \text{cov}(\bar{y}, \bar{x})
\]

\[
\left( \frac{\partial C_{ys}}{\partial \bar{y}} \right) \frac{\partial C_{ys}}{\partial s_{\bar{y}}^2} \left( \frac{\partial C_{ys}}{\partial \bar{y}} \right)_{\theta=\bar{y}} \text{cov}(\bar{y}, s_{\bar{y}}^2) = -\frac{(S_{\bar{y}}^2)^{1/2}}{\bar{Y}^2} \frac{1}{2\bar{Y}(S_{\bar{y}}^2)^{1/2}} \text{cov}(\bar{y}, s_{\bar{y}}^2)
\]

\[
= -\frac{1}{2\bar{Y}^3} \text{cov}(\bar{y}, s_{\bar{y}}^2)
\]

\[
\left( \frac{\partial C_{ys}}{\partial \bar{y}} \right) \frac{\partial C_{ys}}{\partial s_{\bar{y}}^2} \left( \frac{\partial C_{ys}}{\partial \bar{y}} \right)_{\theta=\bar{y}} \text{cov}(\bar{y}, s_{\bar{y}}^2) = \left( -\frac{(S_{\bar{y}}^2)^{1/2}}{\bar{Y}^2} \right) \left( \frac{\gamma}{2\bar{Y}(S_{\bar{y}}^2)^{1/2}} \right) \text{cov}(\bar{y}, s_{\bar{y}}^2)
\]

\[
= -\frac{\gamma}{2\bar{Y}^3} \text{cov}(\bar{y}, s_{\bar{y}}^2)
\]

\[
\left( \frac{\partial C_{ys}}{\partial \bar{y}} \right) \frac{\partial C_{ys}}{\partial \beta} \left( \frac{\partial C_{ys}}{\partial \bar{y}} \right)_{\theta=\bar{y}} \text{cov}(\bar{y}, \beta) = \left( -\frac{(S_{\bar{y}}^2)^{1/2}}{\bar{Y}^2} \right) . 0 . \text{cov}(\bar{y}, \beta) = 0
\]

\[
\left( \frac{\partial C_{ys}}{\partial \bar{y}} \right) \frac{\partial C_{ys}}{\partial s_{\bar{y}}^2} \left( \frac{\partial C_{ys}}{\partial \bar{y}} \right)_{\theta=\bar{y}} \text{cov}(\bar{y}, \bar{y}) = 0
\]

\[
\left( \frac{\partial C_{ys}}{\partial \bar{y}} \right) \frac{\partial C_{ys}}{\partial \bar{y}} \left( \frac{\partial C_{ys}}{\partial \bar{y}} \right)_{\theta=\bar{y}} \text{cov}(\bar{y}, \bar{y}) = 0
\]

\[
\left( \frac{\partial C_{ys}}{\partial \bar{x}} \right) \frac{\partial C_{ys}}{\partial s_{\bar{y}}^2} \left( \frac{\partial C_{ys}}{\partial \bar{x}} \right)_{\theta=\bar{x}} \text{cov}(\bar{x}, s_{\bar{y}}^2) = \left( \beta \frac{(S_{\bar{y}}^2)^{1/2}}{\bar{Y}^2} \right) \left( \frac{1}{2\bar{Y}(S_{\bar{y}}^2)^{1/2}} \right) \text{cov}(\bar{x}, s_{\bar{y}}^2)
\]

\[
= \frac{\beta}{2\bar{Y}^3} \text{cov}(\bar{x}, s_{\bar{y}}^2)
\]

\[
\left( \frac{\partial C_{ys}}{\partial \bar{x}} \right) \frac{\partial C_{ys}}{\partial s_{\bar{y}}^2} \left( \frac{\partial C_{ys}}{\partial \bar{x}} \right)_{\theta=\bar{x}} \text{cov}(\bar{x}, s_{\bar{y}}^2) = \left( \beta \frac{(S_{\bar{y}}^2)^{1/2}}{\bar{Y}^2} \right) \left( \frac{\gamma}{2\bar{Y}(S_{\bar{y}}^2)^{1/2}} \right) \text{cov}(\bar{x}, s_{\bar{y}}^2)
\]

\[
= -\frac{\beta \gamma}{2\bar{Y}^3} \text{cov}(\bar{x}, s_{\bar{y}}^2)
\]
\[
\left(\frac{\partial C_{ys}}{\partial \bar{x}} \right)_{\bar{\theta} = \theta} \left(\frac{\partial C_{ys}}{\partial \hat{\beta}} \right)_{\bar{\theta} = \theta} \text{cov}(\bar{x}, \hat{\beta}) = 0
\]

\[
\left(\frac{\partial C_{ys}}{\partial \bar{y}} \right)_{\bar{\theta} = \theta} \left(\frac{\partial C_{ys}}{\partial \hat{\gamma}} \right)_{\bar{\theta} = \theta} \text{cov}(\bar{x}, \hat{\gamma}) = 0
\]

\[
\left(\frac{\partial C_{ys}}{\partial s_y^2} \frac{\partial C_{ys}}{\partial s_x^2} \right)_{\bar{\theta} = \theta} \text{cov}(s_y^2, s_x^2) = \frac{1}{2\bar{y}(s_y^2)^2} \left( - \frac{1}{2\bar{y}(s_x^2)^2} \right) \text{cov}(s_y^2, s_x^2)
\]

\[
= \frac{\gamma}{4s_y^2s_x^2} \text{cov}(s_y^2, s_x^2)
\]

\[
\left(\frac{\partial C_{ys}}{\partial s_y^2} \frac{\partial C_{ys}}{\partial \hat{\beta}} \right)_{\bar{\theta} = \theta} = \text{cov}(s_y^2, \hat{\beta}) = 0
\]

\[
\left(\frac{\partial C_{ys}}{\partial s_y^2} \frac{\partial C_{ys}}{\partial \hat{\gamma}} \right)_{\bar{\theta} = \theta} = \text{cov}(s_y^2, \hat{\gamma}) = 0
\]

\[
\left(\frac{\partial C_{ys}}{\partial s_x^2} \frac{\partial C_{ys}}{\partial \hat{\beta}} \right)_{\bar{\theta} = \theta} = \text{cov}(s_x^2, \hat{\beta}) = 0
\]

\[
\left(\frac{\partial C_{ys}}{\partial s_x^2} \frac{\partial C_{ys}}{\partial \hat{\gamma}} \right)_{\bar{\theta} = \theta} = \text{cov}(s_x^2, \hat{\gamma}) = 0
\]

\[
\left(\frac{\partial C_{ys}}{\partial \hat{\beta}} \frac{\partial C_{ys}}{\partial \hat{\gamma}} \right)_{\bar{\theta} = \theta} = \text{cov}(\hat{\beta}, \hat{\gamma}) = 0
\]

Inserting all the above derivates in Result 1.2, after some elementary algebra this leads to

\[
\text{Bias}(\hat{c}_{ys}) = \beta \frac{s_y^2}{\bar{y}^2} \text{cov}(\bar{y}, \bar{x}) - \frac{1}{2\bar{y}^3} \text{cov}(\bar{y}, s_y^2) + \frac{\gamma}{2\bar{y}^3} \text{cov}(\bar{y}, s_x^2) + \frac{\beta}{2\bar{y}^3} \text{cov}(\bar{x}, s_y^2)
\]

\[
+ \frac{\beta \gamma}{2\bar{y}^3} \text{cov}(\bar{x}, s_x^2) - \frac{\gamma}{4\bar{y}^2(s_y^2)^2} \text{cov}(s_y^2, s_x^2)
\]

\[\text{(5.19)}\]
\begin{align*}
V(\hat{C}_{ys}) &= \frac{s_y^2}{Y^4} v(y) + \beta^2 \frac{s_y^2}{Y^4} v(x) + \frac{1}{4Y^2 s_y^2} v(s_y^2) + \frac{\gamma^2}{4Y^2 s_y^2} V(s_x^2) - 2\beta \frac{s_y^2}{Y^4} \text{cov}(y, x) \\
&\quad - \frac{1}{\bar{Y}^3} \text{cov}(\bar{Y}, s_y^2) + \frac{\gamma}{\bar{Y}^3} \text{cov}(\bar{Y}, s_x^2) + \frac{\beta}{\bar{Y}^3} \text{cov}(x, s_y^2) - \frac{\beta \gamma}{\bar{Y}^3} \text{cov}(x, s_x^2) \\
&\quad - \frac{\gamma}{2\bar{Y}^2 s_y^2} \text{cov}(s_y^2, s_x^2) \tag{5.20}
\end{align*}

Now, inserting \( \beta = \frac{\text{cov}(\bar{Y}, x)}{v(x)} \) and \( \gamma = \frac{\text{cov}(s_y^2, s_x^2)}{v(s_x^2)} \) in (5.19) and (5.20) and then using (5.18) (5.19) and (5.20) can be rewritten as

\begin{align*}
\text{Bias}(\hat{C}_{ys}) &= \frac{1-f}{n} C_y^2 \left[ -\rho^2 C_y^2 - \frac{1}{2} \rho C_y \lambda_30 + \frac{1}{2} C_y \frac{\lambda_{22} \lambda_{12}}{\lambda_{04}} + \frac{1}{2} \rho C_y \lambda_{21} - \frac{1}{2} \rho C_y \frac{\lambda_{22} \lambda_{30}}{\lambda_{04}} \right. \\
&\quad \left. - \frac{1}{4} \frac{\lambda_{22}}{\lambda_{04}} \right] \tag{5.21}
\end{align*}

and

\begin{align*}
V(\hat{C}_{ys}) &\approx \frac{1-f}{n} C_y^2 \left[ C_y^2 - \rho^2 C_y^2 + \frac{1}{4} C_y \lambda_{40} - \frac{1}{4} \frac{\lambda_{22}}{\lambda_{04}} - 2\rho^2 C_y^2 - C_y \lambda_{30} + C_y \frac{\lambda_{22} \lambda_{12}}{\lambda_{04}} \\
&\quad + \rho C_y \lambda_{21} - \frac{1}{2} \rho C_y \frac{\lambda_{22} \lambda_{30}}{\lambda_{04}} - \frac{1}{2} \frac{\lambda_{22}}{\lambda_{04}} \right] \tag{5.22}
\end{align*}

**Remark 5.1.** Since the ratio-type, optimal type, calibration-type and Isaki-type estimators are special cases of the system of estimators \( \hat{C}_{ys} \), their variance expression can be deduced from (5.22) with suitable modifications.
5.4 Simulation Study

We now turn to the comparison of relative efficiency of the suggested estimators with that of the conventional estimator under SRSWOR. These estimators make use of the auxiliary information at the estimation stage. For the comparison, we used the populations listed in Appendix 5.A.

From each of populations 1 to 11 a sample of size \( n = 30 \), whereas from each of populations 11 to 19 a sample of size 15 was drawn using SRSWOR. The estimators \( \hat{\epsilon}_y, \hat{\epsilon}_{yr}, \hat{\epsilon}_{yo}, \hat{\epsilon}_{yc} \), and \( \hat{\epsilon}_{yl} \) were computed from each sample. This process was repeated \( M = 10,000 \) times. The performance of the different estimators was measured and compared in terms of percentage relative bias (\% RB) and relative efficiency (RE). These summary statistics over 10,000 samples were calculated for a particular estimator \( \hat{\epsilon} \) as

\[
\%RB(\hat{\epsilon}) = 100 \times \frac{\hat{\epsilon} - C_y}{C_y} \quad \text{and} \quad RE(\hat{\epsilon}) = 100 \times \frac{MSE(\hat{\epsilon})}{MSE(C_y)}
\]

where \( \bar{\hat{\epsilon}} = \frac{1}{M} \sum_{j=1}^{M} \hat{\epsilon}^{(j)} \), and \( MSE(\hat{\epsilon}) = \frac{1}{M-1} \sum_{j=1}^{M} (\hat{\epsilon}^{(j)} - C_y)^2 \)

Tables 5.1 report the RB\% and RE for the estimators included in the simulation study.
### Table 5.1  RB (%) and RE under SRSWOR.

<table>
<thead>
<tr>
<th>pop.</th>
<th>(\hat{C}_y)</th>
<th>(\hat{C}_{yo})</th>
<th>(\hat{C}_{yr})</th>
<th>(\hat{C}_{yc})</th>
<th>(\hat{C}_{yl})</th>
<th>(\hat{C}_y)</th>
<th>(\hat{C}_{yo})</th>
<th>(\hat{C}_{yr})</th>
<th>(\hat{C}_{yc})</th>
<th>(\hat{C}_{yl})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-4.282</td>
<td>-1.858</td>
<td>-2.081</td>
<td>-2.797</td>
<td>-2.90</td>
<td>100</td>
<td>84</td>
<td>84</td>
<td>86</td>
<td>83</td>
</tr>
<tr>
<td>2</td>
<td>-4.282</td>
<td>-1.137</td>
<td>0.6488</td>
<td>-0.558</td>
<td>-1.39</td>
<td>100</td>
<td>314</td>
<td>351</td>
<td>456</td>
<td>375</td>
</tr>
<tr>
<td>3</td>
<td>-0.259</td>
<td>0.508</td>
<td>0.611</td>
<td>0.294</td>
<td>0.308</td>
<td>100</td>
<td>271</td>
<td>253</td>
<td>255</td>
<td>284</td>
</tr>
<tr>
<td>4</td>
<td>-6.531</td>
<td>-2.742</td>
<td>-2.022</td>
<td>-2.191</td>
<td>-2.14</td>
<td>100</td>
<td>481</td>
<td>595</td>
<td>561</td>
<td>521</td>
</tr>
<tr>
<td>5</td>
<td>-9.106</td>
<td>-2.610</td>
<td>-1.977</td>
<td>-2.477</td>
<td>-2.04</td>
<td>100</td>
<td>925</td>
<td>1008</td>
<td>984</td>
<td>924</td>
</tr>
<tr>
<td>6</td>
<td>-6.679</td>
<td>0.660</td>
<td>0.828</td>
<td>0.709</td>
<td>0.709</td>
<td>100</td>
<td>5928</td>
<td>7828</td>
<td>6861</td>
<td>8079</td>
</tr>
<tr>
<td>7</td>
<td>-0.953</td>
<td>-0.800</td>
<td>-1.095</td>
<td>-0.899</td>
<td>-1.11</td>
<td>100</td>
<td>214</td>
<td>248</td>
<td>213</td>
<td>206</td>
</tr>
<tr>
<td>8</td>
<td>-14.4</td>
<td>-12.35</td>
<td>-13.44</td>
<td>-13.11</td>
<td>-13.3</td>
<td>100</td>
<td>90</td>
<td>104</td>
<td>98</td>
<td>96</td>
</tr>
<tr>
<td>9</td>
<td>-0.275</td>
<td>0.774</td>
<td>1.036</td>
<td>0.607</td>
<td>0.22</td>
<td>100</td>
<td>99</td>
<td>74</td>
<td>98</td>
<td>106</td>
</tr>
<tr>
<td>10</td>
<td>-0.345</td>
<td>0.600</td>
<td>0.805</td>
<td>0.414</td>
<td>0.07</td>
<td>100</td>
<td>101</td>
<td>75</td>
<td>98</td>
<td>107</td>
</tr>
<tr>
<td>11</td>
<td>-0.209</td>
<td>0.245</td>
<td>0.111</td>
<td>0.371</td>
<td>0.31</td>
<td>100</td>
<td>674</td>
<td>589</td>
<td>316</td>
<td>646</td>
</tr>
<tr>
<td>12</td>
<td>-1.167</td>
<td>-0.976</td>
<td>-0.674</td>
<td>-0.677</td>
<td>-0.61</td>
<td>100</td>
<td>585</td>
<td>608</td>
<td>609</td>
<td>594</td>
</tr>
<tr>
<td>13</td>
<td>-7.903</td>
<td>-3.514</td>
<td>-2.976</td>
<td>-2.865</td>
<td>-2.59</td>
<td>100</td>
<td>830</td>
<td>859</td>
<td>901</td>
<td>819</td>
</tr>
<tr>
<td>14</td>
<td>-7.613</td>
<td>-2.665</td>
<td>-2.379</td>
<td>-2.204</td>
<td>-1.98</td>
<td>100</td>
<td>1736</td>
<td>1705</td>
<td>1755</td>
<td>1605</td>
</tr>
<tr>
<td>15</td>
<td>-8.385</td>
<td>-2.056</td>
<td>-1.961</td>
<td>-1.734</td>
<td>-1.61</td>
<td>100</td>
<td>3259</td>
<td>2998</td>
<td>3266</td>
<td>3015</td>
</tr>
<tr>
<td>16</td>
<td>-8.394</td>
<td>-0.406</td>
<td>-0.513</td>
<td>-0.452</td>
<td>-0.45</td>
<td>100</td>
<td>6048</td>
<td>7276</td>
<td>8388</td>
<td>8455</td>
</tr>
<tr>
<td>17</td>
<td>-7.947</td>
<td>-3.903</td>
<td>-3.583</td>
<td>-3.168</td>
<td>-2.91</td>
<td>100</td>
<td>1275</td>
<td>1150</td>
<td>1139</td>
<td>1010</td>
</tr>
<tr>
<td>18</td>
<td>0.692</td>
<td>0.332</td>
<td>0.243</td>
<td>0.252</td>
<td>0.218</td>
<td>100</td>
<td>9889</td>
<td>24360</td>
<td>16757</td>
<td>18026</td>
</tr>
<tr>
<td>19</td>
<td>-0.051</td>
<td>0.259</td>
<td>0.084</td>
<td>0.153</td>
<td>0.103</td>
<td>100</td>
<td>159.013</td>
<td>38750</td>
<td>33953</td>
<td>37897</td>
</tr>
</tbody>
</table>

Tables 5.1 and 5.2 suggest the following comments.

- The absolute values of RBs of all the suggested estimators are within ± 4 % (except population 8) compared to the convention estimator. The convention estimator has considerably underestimated the true CV for most of the populations under study.
The ratio-type, optimal-type, calibration and Isaki-type estimators have performed very well (except the populations 1, 8 and 9, the reason might be that the fraction of the total variation explained by the regression line is low).

The REs of suggested estimators increase with increasing the correlation between x and y when CV(x) is low (see population 11) or when it is above 1.

5.5 Conclusion

In many survey populations strong auxiliary information is available and the relationship between y and x is often a straight line through the origin (e.g., see, Royall and Cumberland 1981a, b). Exploiting this relationship we have suggested various estimators of finite population CV, which is an important parameter of interest for future study.

The conclusions emerging from the simulation study can be summarized as follow.

- Implementation of the suggested estimators requires the complete auxiliary information. It is clear that \( \hat{C}_{yr}, \hat{C}_{yo}, \hat{C}_{yc}, \) and \( \hat{C}_{yl} \) will reflect the true CV closely when the best linear fit goes through the origin and residuals from it are small.

- With strong auxiliary information the gains from using the suggested estimators are substantial as compared to customary estimator. However, overall \( \hat{C}_{yl} \) is slightly better than the other three with respect to both criteria.
## Appendix 5.A (List of Study populations)

<table>
<thead>
<tr>
<th>Pop</th>
<th>N</th>
<th>Source</th>
<th>X</th>
<th>Y</th>
<th>(\rho_{xy})</th>
<th>CV(y)</th>
<th>CV(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>281</td>
<td>Särndal et al. (1992)</td>
<td>CS82: Number of conservative seats in municipal council</td>
<td>RMT \times 10^{-4}: Revenue from the 1985 municipal taxation</td>
<td>0.657</td>
<td>1.058</td>
<td>0.519</td>
</tr>
<tr>
<td>2</td>
<td>338</td>
<td>Chambers et al. (1986)</td>
<td>Area assigned for sugarcane farms</td>
<td>Gross value of sugarcane</td>
<td>0.902</td>
<td>0.610</td>
<td>0.519</td>
</tr>
<tr>
<td>3</td>
<td>393</td>
<td>Valliant et al. (2000)</td>
<td>Number of beds</td>
<td>Number of patients discharged</td>
<td>0.910</td>
<td>0.724</td>
<td>0.776</td>
</tr>
<tr>
<td>4</td>
<td>301</td>
<td>Valliant et al. (2000)</td>
<td>Adult female population, 1960</td>
<td>Breast cancer mortality, 1950-69 (white female)</td>
<td>0.967</td>
<td>1.279</td>
<td>1.221</td>
</tr>
<tr>
<td>5</td>
<td>304</td>
<td>Valliant et al. (2000)</td>
<td>Number of households 1960</td>
<td>Population, excluding residents of group quarters, 1960</td>
<td>0.982</td>
<td>1.380</td>
<td>1.302</td>
</tr>
<tr>
<td>6</td>
<td>304</td>
<td>Valliant et al. (2000)</td>
<td>Number of households 1960</td>
<td>Population, excluding Residents of group quarters, 1960</td>
<td>0.998</td>
<td>1.239</td>
<td>1.302</td>
</tr>
<tr>
<td>7</td>
<td>396</td>
<td>Platt et al. (deleted last three observations)</td>
<td>The diameter in centimeters at breast height</td>
<td>The entire height in feet</td>
<td>0.907</td>
<td>1.084</td>
<td>0.8402</td>
</tr>
<tr>
<td>8</td>
<td>128</td>
<td>Murthy(1967) p.127-130</td>
<td>Number of persons</td>
<td>Workers at household industry</td>
<td>0.540</td>
<td>1.347</td>
<td>0.605</td>
</tr>
<tr>
<td>9</td>
<td>128</td>
<td>Murthy(1967) p.127-130</td>
<td>Cultivated area (in acres)</td>
<td>No. of cultivators</td>
<td>0.567</td>
<td>0.673</td>
<td>0.572</td>
</tr>
<tr>
<td>10</td>
<td>128</td>
<td>Murthy(1967) p.127-130</td>
<td>Number of persons</td>
<td>No. of cultivators</td>
<td>0.831</td>
<td>0.673</td>
<td>0.605</td>
</tr>
<tr>
<td>11</td>
<td>80</td>
<td>Murthy(1967) p.228</td>
<td>Fixed capital</td>
<td>Output for factories</td>
<td>0.941</td>
<td>0.751</td>
<td>0.354</td>
</tr>
<tr>
<td>12</td>
<td>34</td>
<td>Murthy(1967) p.399</td>
<td>Area under wheat in 1963</td>
<td>Area under wheat in 1964</td>
<td>0.980</td>
<td>0.753</td>
<td>0.721</td>
</tr>
<tr>
<td>13</td>
<td>51</td>
<td>SPSS 14.0 map data</td>
<td>Total population in 1990</td>
<td>Total population in 1998</td>
<td>0.998</td>
<td>1.116</td>
<td>1.115</td>
</tr>
<tr>
<td>14</td>
<td>51</td>
<td>SPSS 14.0 map data</td>
<td>Total households in 1990</td>
<td>Total population in 1998</td>
<td>0.999</td>
<td>1.116</td>
<td>1.077</td>
</tr>
<tr>
<td>15</td>
<td>51</td>
<td>SPSS 14.0 map data</td>
<td>Total households in 1998</td>
<td>Total population in 1998</td>
<td>0.999</td>
<td>1.116</td>
<td>1.077</td>
</tr>
<tr>
<td>16</td>
<td>51</td>
<td>SPSS 14.0 map data</td>
<td>Total households in 1990</td>
<td>Total population in 1990</td>
<td>0.999</td>
<td>1.115</td>
<td>1.087</td>
</tr>
<tr>
<td>17</td>
<td>51</td>
<td>SPSS 14.0 map data</td>
<td>Total households in 1990</td>
<td>Total population in 1990</td>
<td>0.999</td>
<td>1.115</td>
<td>1.087</td>
</tr>
<tr>
<td>18</td>
<td>31</td>
<td><a href="http://www.india-stat.com">www.india-stat.com</a></td>
<td>Female population in 1971</td>
<td>Female population in 1971</td>
<td>1.000</td>
<td>1.245</td>
<td>1.250</td>
</tr>
<tr>
<td>19</td>
<td>31</td>
<td><a href="http://www.india-stat.com">www.india-stat.com</a></td>
<td>Male population in 1971</td>
<td>Male population in 1971</td>
<td>1.000</td>
<td>1.262</td>
<td>1.260</td>
</tr>
</tbody>
</table>

113
(Graphs of the populations listed in Appendix 5.A)