

# A Family of Estimators for Estimating Population Variance Using Auxiliary Information in Sample Survey

Hilal A. Lone  
Vikram University Ujjain (M.P.) India  
hilalstat@gmail.com

Rajesh Tailor  
Vikram University Ujjain (M.P.) India  
tailorraj@gmail.com

## Abstract

In this article an efficient class of estimators for estimating finite population variance has been proposed using auxiliary information in simple random sampling. The bias and mean squared error of the proposed estimator is obtained up to the first degree of approximation. It has been shown that the proposed estimator is more efficient than usual unbiased estimator, Isaki (J. Am. Stat. Assoc. 78:117-123, 1983), Kadilar and Cingi (Appl. Math. & Comput., 173, 1047-1059, 2006) and Upadhyaya and Singh (Vikram Math. J. 19, 14-17, 1999a). To judge the merits of the proposed estimator, we consider one numerical example.

**Keywords:** Mean squared error, Bias, Auxiliary information, Family of estimators.

## 1. Introduction

In survey sampling an attempts have been made by various researchers to improve the efficiency of the estimators by using the auxiliary information. In this context ratio, product and ratio-cum-product estimators are good examples. When coefficient of variation  $c_y$  of the study variate  $y$  is known, Searl (1964) considered the problem of estimating population mean  $\bar{Y}$ . Motivated by Searl (1964), Sisodia and Dwivedi (1981) used coefficient of variation  $C_x$  of the auxiliary variate  $x$ . Later many authors utilized this information. Singh et al. (2004) proposed ratio and product type estimators using coefficient of kurtosis  $\beta_2(x)$  of auxiliary variate  $x$  whereas Upadhyaya and Singh (1999b) utilized both information coefficient of variation  $C_x$  as well as coefficient of kurtosis  $\beta_2(x)$  of the auxiliary variate  $x$ . Latter the problem of estimating finite population mean has been discussed by various researchers including Singh (1967), Panday and Dubey (1988), Singh and Biradar (2002), Singh and Ruiz Espejo (2003), Tailor and Tailor (2008), Sharma and Tailor (2010), Tailor et al. (2011), Solanki and Singh (2013) and Tailor and Lone (2014).

The problem of estimating population variance has also attracted the attentions of researchers in survey sampling. Das and Tripathi (1978) have discussed the problem of variance estimation under the situations of known population variance and coefficient of variation of the auxiliary variate. Das (1988) has proposed some wider classes of estimators for estimating finite population variance. Isaki (1983), Kadilar and Cingi (2006), Singh and Chandra (2008), Dubey and Sharma (2008), Gupta and Shabbir (2008), Singh and Solanki (2013) and Tailor and Lone (2013) and others contributed well in estimating the finite population variance.

Let  $U = \{U_1, U_2, \dots, U_N\}$  be a finite population of  $N$  units. Let  $y$  be the study variate and  $x$  be the auxiliary variates observed on  $U_i$  ( $i=1,2,3,\dots,N$ ), where  $x$  is highly correlated with the study variate  $y$ . A sample of size  $n$  is drawn from population  $U$  using simple random sampling without replacement.

It is also assumed that the population size  $N$  is very large so that the finite population correction (FPC) term is ignored.

Let us define:

$$s_y^2 = S_y^2(1 + e_0), \quad s_x^2 = S_x^2(1 + e_1) \quad \text{and} \quad \bar{x} = \bar{X}(1 + e_2) \quad \text{such that}$$

$$E(e_0) = E(e_1) = E(e_2) = 0,$$

$$E(e_0^2) = \frac{1}{n}(\lambda_{40} - 1) = n^{-1}\lambda_{40}^*, \quad E(e_2^2) = \frac{1}{n}C_x^2, \quad E(e_1^2) = \frac{1}{n}(\lambda_{04} - 1) = n^{-1}\lambda_{04}^*,$$

$$E(e_0e_2) = \frac{1}{n}\lambda_{21}C_x, \quad E(e_0e_1) = \frac{1}{n}(\lambda_{22} - 1) = n^{-1}\lambda_{22}^*, \quad \text{and} \quad E(e_1e_2) = \frac{1}{n}C_x\lambda_{03}.$$

where  $\lambda_{pq} = \frac{\mu_{pq}}{(\mu_{20}^{p/2}\mu_{02}^{q/2})}$  and  $\mu_{pq} = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{Y})^p (x_i - \bar{X})^q$ ;  $(p, q)$  being non negative integers.

## 2. Procedure, Notations and Definitions

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 : \text{Sample variance of the study variate } y$$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 : \text{Sample variance of the study variate } x$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i : \text{Unbiased estimator of population mean } \bar{X}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i : \text{Unbiased estimator of population mean } \bar{Y}$$

$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})^2 : \text{Population variance of the auxiliary variate } x$$

$$S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})^2 : \text{Population variance of the study variate } y$$

$$S_{yx} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y}) : \text{Population covariance of } x \text{ and } y$$

Isaki (1983) proposed the ratio estimator for  $S_y^2$  as

$$t_1 = s_y^2 \left( \frac{S_x^2}{s_x^2} \right), \quad (2.1)$$

Kadilar and Cingi (2006) considered the following ratio type estimators for  $S_y^2$  as

$$t_2 = s_y^2 \left( \frac{S_x^2 - C_x}{s_x^2 - C_x} \right), \quad (2.2)$$

$$t_3 = s_y^2 \left( \frac{S_x^2 - \beta_2(x)}{s_x^2 - \beta_2(x)} \right), \quad (2.3)$$

$$t_4 = s_y^2 \left( \frac{S_x^2 \beta_2(x) - C_x}{s_x^2 \beta_2(x) - C_x} \right), \quad (2.4)$$

$$t_5 = s_y^2 \left( \frac{C_x S_x^2 - \beta_2(x)}{C_x s_x^2 - \beta_2(x)} \right), \quad (2.5)$$

Upadhyaya and Singh (1999a) proposed ratio estimator for  $S_y^2$  as

$$t_6 = s_y^2 \left( \frac{S_x^2 + \beta_2(x)}{s_x^2 + \beta_2(x)} \right), \quad (2.6)$$

The mean squared error of the estimators  $t_i$  ( $i = 0, 1, 2, 3, \dots, 6$ ) up to the first degree of approximation are given as

$$MSE(t_i) = S_y^4 n^{-1} \left[ \lambda_{40}^* + \delta_i \lambda_{04}^* (\delta_i - 2K) \right], \quad (i = 0, 1, 2, 3, \dots, 6) \quad (2.10)$$

where  $K = \lambda_{22}^* \lambda_{04}^{*-1}$

$$\delta_i = \begin{cases} 0, & i = 0 \\ 1, & i = 1 \\ S_x^2 / (S_x^2 - C_x), & i = 2 \\ S_x^2 / (S_x^2 - \beta_2(x)), & i = 3 \\ S_x^2 \beta_2(x) / (S_x^2 \beta_2(x) - C_x), & i = 4 \\ S_x^2 C_x / (S_x^2 C_x - \beta_2(x)), & i = 5 \\ S_x^2 / (S_x^2 + \beta_2(x)), & i = 6 \end{cases} \quad (2.11)$$

### 3. The Proposed Class of Estimators

We define the following class of estimators for the population variance  $S_y^2$  as

$$H = \left[ W_1 s_y^2 \left( \frac{aS_x^2 - b}{as_x^2 - b} \right) + W_2 s_y^2 \left( \frac{\theta \bar{x} - \varphi}{\theta \bar{X} - \varphi} \right) \right] \quad (3.1)$$

It is to be noted that  $(W_1, W_2)$  are suitably chosen constant, can be determined such that mean squared error of the estimator  $H$  is minimum and  $(a, b, \theta, \varphi)$  are either constants or function of known parameters  $C_x$ ,  $\beta_2(x)$  and  $\rho_{yx}$  of the auxiliary variate  $x$ . We would like to mention that for different values of  $(a, b, \theta, \varphi)$ , we get seven estimators as shown in table 3.1. Note that in table 3.1  $t_1$  is the estimator proposed by Isaki (1983),  $t_i$   $i = (2, 3, 4, 5)$  are the estimators proposed by Kadilar and Cinghi (2006),  $t_6$  is the estimator proposed by Upadhyaya and Singh 1999a).

Expressing (3.1) in terms of  $e$ 's, we have

$$H = S_y^2 [W_1 + W_2 - W_1 M e_1 + N W_2 e_2 + W_1 M^2 e_1^2 + W_1 e_0 + W_2 e_0 - W_1 M e_0 e_1 + W_2 N e_0 e_2] \\ (H - S_y^2) = S_y^2 [W_1 (1 + e_0 - M e_1 - M e_0 e_1 + M^2 e_1^2) + W_2 (1 + e_0 + N e_2 + N e_0 e_2) - 1] \quad (3.2)$$

Taking expectation on both sides to (3.2), we get the bias of the estimator  $H$  up to the first degree of approximation as

$$\text{Bias}(H) = S_y^2 [A W_1 + B W_2 - 1] \quad (3.3)$$

where  $A = 1 + M^2 n^{-1} \lambda_{04}^* - M n^{-1} \lambda_{22}^*$

$$B = 1 + S n^{-1} \lambda_{22}^*$$

Squaring and taking expectations on both sides to (3.2), we get the mean squared error of the estimator  $H$ , up to the first degree of approximation as

$$\text{MSE}(H) = S_y^4 [1 + C W_1^2 + D W_2^2 + 2 E W_1 W_2 - 2 W_1 F - 2 W_2 G] \quad (3.4)$$

where

$$C = 1 + 3 M^2 n^{-1} \lambda_{04}^* + n^{-1} \lambda_{40}^* - 4 M n^{-1} \lambda_{22}^*$$

$$D = 1 + n^{-1} \lambda_{40}^* + S^2 n^{-1} C_x^2 + 4 S n^{-1} C_x \lambda_{21}$$

$$E = 1 + 2 S n^{-1} C_x \lambda_{21} - 2 M n^{-1} \lambda_{22}^* - M S n^{-1} C_x \lambda_{03} + n^{-1} \lambda_{40}^* + M^2 n^{-1} \lambda_{04}^*$$

$$F = 1 + M^2 n^{-1} \lambda_{04}^* - M n^{-1} \lambda_{22}^*$$

$$G = 1 + S n^{-1} C_x \lambda_{21}$$

The mean squared error of the estimators  $H$  is minimized when

$$\left. \begin{aligned} W_1 &= \frac{(DF - EG)}{(CD - E^2)} = W_1^{opt} \\ W_2 &= \frac{(CG - EF)}{(CD - E^2)} = W_2^{opt} \end{aligned} \right\} \quad (3.5)$$

Putting (3.5) in (3.4), we get the minimum mean squared error of the estimator  $H$  as

$$\text{Min. MSE}(H) = S_y^4 \left[ 1 - \frac{(DF^2 + CG^2 - 2EFG)}{(CD - E^2)} \right] \quad (3.6)$$

$$\text{Min. MSE}(H) = S_y^4 [1 - K^*] \quad (3.7)$$

$$\text{where } K^* = \frac{(DF^2 + CG^2 - 2EFG)}{(CD - E^2)}$$

where  $C, D, E$  and  $F$  have their usual meanings.

**Table 3.1: Some known members of  $H$**

VALUES OF CONSTANTS						ESTIMATORS
$a$	$b$	$\theta$	$\phi$	$W_1$	$W_2$	
0	1	-	-	1	0	$t_0$ (Usual unbiased estimator)
0	1	-	-	1	0	$t_1$ (Isaki 1983)
1	$C_x$	-	-	1	0	$t_2$ (Kadilar and Cingi 2006)
1	$\beta_2(x)$	-	-	1	0	$t_3$ (Kadilar and Cingi 2006)
$\beta_2(x)$	$C_x$	-	-	1	0	$t_4$ (Kadilar and Cingi 2006)
$C_x$	$\beta_2(x)$	-	-	1	0	$t_5$ (Kadilar and Cingi 2006)
1	$-\beta_2(x)$	-	-	1	0	$t_6$ (Upadhyaya and Singh 1999)

## 6. Efficiency comparisons of the estimator $H$ with the estimators $t_i$ ( $i=0, 1, 2, \dots, 6$ )

From (3.7) and (2.10), it is observed that the proposed estimator would be more efficient than

(i) Usual unbiased estimator  $S_y^2$  if

$$K^* > 1 - n^{-1} \lambda_{40}^* \quad (6.1)$$

(ii) Isaki (1983) estimator  $t_1$  if

$$K^* + 2K n^{-1} \lambda_{04}^* > 1 - n^{-1} [\lambda_{40}^* + \lambda_{04}^*] \quad (6.2)$$

(iii) Kadilar and Cingi (2006) estimators  $t_2$  if

$$K^* + 2K n^{-1} \lambda_{04}^* S_x^2 / (S_x^2 - C_x) > 1 - n^{-1} [\lambda_{40}^* + (S_x^2 / (S_x^2 - C_x))^2 \lambda_{04}^*] \quad (6.3)$$

(iii) Kadilar and Cingi (2006) estimators  $t_3$  if

$$K^* + 2K n^{-1} \lambda_{04}^* S_x^2 / (S_x^2 - \beta_2(x)) > 1 - n^{-1} \left[ \lambda_{40}^* + (S_x^2 / (S_x^2 - \beta_2(x)))^2 \lambda_{04}^* \right] \quad (6.4)$$

(iv) Kadilar and Cingi (2006) estimators  $t_4$  if

$$K^* + 2K n^{-1} \lambda_{04}^* S_x^2 \beta_2(x) / (S_x^2 \beta_2(x) - C_x) > 1 - n^{-1} \left[ \lambda_{40}^* + (S_x^2 \beta_2(x) / (S_x^2 \beta_2(x) - C_x))^2 \lambda_{04}^* \right] \quad (6.5)$$

(v) Kadilar and Cingi(2006) estimators  $t_5$  if

$$K^* + 2K n^{-1} \lambda_{04}^* S_x^2 C_x / (S_x^2 C_x - \beta_2(x)) > 1 - n^{-1} \left[ \lambda_{40}^* + (S_x^2 C_x / (S_x^2 C_x - \beta_2(x)))^2 \lambda_{04}^* \right] \quad (6.6)$$

(vi) Upadhyaya and Singh (1999a)  $t_6$  if

$$K^* + 2K n^{-1} \lambda_{04}^* S_x^2 / (S_x^2 + \beta_2(x)) > 1 - n^{-1} \left[ \lambda_{40}^* + (S_x^2 / (S_x^2 + \beta_2(x)))^2 \lambda_{04}^* \right] \quad (6.7)$$

$$\text{where } K^* = \frac{(DF^2 + CG^2 - 2EFG)}{(CD - E^2)}$$

It follows from (6.1) to (6.7) that the proposed class of estimator is more efficient than usual unbiased estimator, Isaki (1983) estimator, Kadilar and Cingi (2006) estimators and Upadhyaya and Singh (1999a) estimator.

## 7. Empirical Study

To exhibit the performance of the suggested class of estimators in comparison to other estimators, we consider a natural population from [Singh (2003), p.1111-1112]. The description of population is given below

$y$ : Amount (in \$000) of real estate farm loans in different states during 1997,  $x$ : Amount (in \$000) of non-real estate farm loans in different states during 1997.

**Table 7.1**

---

$\lambda_{40}=3.5822$ ,  $\lambda_{04}=4.5247$ ,  $\lambda_{22}=2.8411$ ,  $\lambda_{21}=0.9387$ ,  $\lambda_{12}=1.0982$ ,  
 $\lambda_{03}=1.5936$ ,  $\bar{Y}=555.43$ ,  $\bar{X}=878.16$ ,  $C_x=1.2351$ ,  $C_y=1.0529$ ,  $n=10$

---

**Percent Relatives Efficiencies of  $S_y^2$ ,  $t_i$  ( $i=1, 2, \dots, 6$ ) and  $H$  with respect to  $S_y^2$**

Estimators	PRE	Estimators	PRE
$t_0$	100.00	$t_4$	156.0172
$t_1$	156.0173	$t_5$	156.0176
$t_2$	156.0157	$t_6$	156.0179
$t_3$	156.0168	$H$	<b>163.8827</b>

---

## 8. Conclusion

In table 7.2, it is observed that the proposed estimator is more efficient than usual unbiased estimator, Isaki (1983) estimator, Kadilar and Cingi (2006) estimators and Upadhaya and Singh (1999a) estimator. Section 6 deals with the theoretical efficiency comparisons of considered estimators and provided the conditions under which the proposed estimator  $H$  has less mean squared error in comparisons to usual unbiased estimator  $t_0$ , Isaki (1983) estimator  $t_1$ , Kadilar and Cingi (2006) estimators  $t_i$  ( $i = 2,3,4,5$ ) and Upadhaya and Singh (1999a) estimator  $t_6$ . Thus the proposed estimator is recommended for use in practice if the conditions defined in the section 6 are satisfied.

## References

1. Das, A.K. (1988). Contributions to the theory of sampling strategies based on auxiliary information. Unpublished Ph.D thesis, B.C.K.V., Mohanpur, Nadia, West Bengal.
2. Das, A.K. and Tripathi, T.P. (1978). Use of auxiliary information in estimating the finite population variance. *Sankhya*, C. 40, 139-148.
3. Dubey, V. and Sharama, H.K. (2008). On estimation population variance using auxiliary information. *Statist. in Transi.-new Series*, 9,1,7-18.
4. Gupta, S. and Shabbir, J. (2008). Variance estimation in simple random sample using auxiliary information. *Haceteppe J. of Math. & Statist.*, 37(1), 57-67.
5. Isaki, C.T. (1983). Variance estimation using auxiliary information. *J.Amer.Statist.Assoc.* 78, 117-123.
6. Kadilar, C. and Cingi, H. (2006). Ratio estimators for the population in simple and stratified random sampling. *Applied Maths. & Comp.*, 173, 1047-1059.
7. Pandey, B. N. and Dubey, V. (1988). Modified product estimator using coefficient of variation of auxiliary variable. *Assam Statist. Rev.*, 2, 64-66.
8. S. Singh (2003). Advanced sampling theory with applications, Kluwer Academic Press.
9. Searl, D. T. (1964). The utilization of a known coefficient of variation in the estimation procedure. *J. Amer. Statist. Assoc.*, 59, 1225-1226.
10. Sharma, B.K. and Tailor, R. (2010). A new ratio-cum-dual to ratio estimator of finite population mean in simple random sampling. *Global J. of Sci. Frontier Research*, 10, 1(ver1.0), 27-31.
11. Singh, H.P. and Chandra, P. (2008). An alternative to ratio estimator of the population variance in sample surveys. *Statist. in Transi.*, 9, 1, 89-103.
12. Singh, H.P. and Ruiz Espejo, M. (2003). On Linear regression and ratio-product estimation of a finite population mean. *The Statistician*, 52,1,59-67.
13. Singh, M.P. (1967). Ratio-cum-product method of estimation. *Metrika*, 12, 34-42.
14. Singh, H.P. and Biradar, R.S. (2002). A family of alternative to ratio-cum-product estimator in sample surveys. *Assam Statist. Review*, 16,1,81-89.

15. Singh, H.P. and Solanki, R.S. (2013). A new procedure for variance estimation in simple random sampling using auxiliary information. *Statist. Pap.*, 54, 2, 479-497.
16. Sisodia, B. V. S. and Dwivedi, V. K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. *J. Indian Soc. Agri. Statist.*, 33, 1, 13-18.
17. Solanki, R.S. and Singh, H.P. (2013). Efficient classes of estimators in stratified random Sampling. *Statist. Papers*, DOI 10.1007/s00362-013-0567-1.
18. Tailor R. and Lone, H. A. (2014). Ratio –cum- product estimator of finite population mean in double sampling for stratification. *J. of Reliability & Statist. Studies.*, 7, 1, 93-101.
19. Tailor, R., Sharma, B.K. and Kim, J. M. (2011). A generalized ratio-cum-product estimator of finite population mean in stratified random sampling. *Comm. Korean Statist. Soc.* 18, 1,111-118.
20. Tailor, R. and Lone, H. A. (2013). Ratio-cum-product type estimator for population variance in sample surveys. *Octagon Mathematical J.*, 21, 2, pp. 649-654.
21. Tailor, R. and Tailor, R. (2008). Estimation of finite population mean using two auxiliary variables in sample surveys. *Model Assist. Statist. Appl.*, 3, 4, 297-303.
22. Upadhyaya, L. N. and Singh, H. P. (1999a). An estimator for finite population variance that utilizes the kurtosis of an auxiliary variable in sample survey. *Vikram Math. J.* 19, 14-17.
23. Upadhyaya, L. N. and Singh, H. P. (1999b). Use of transformed auxiliary variable in estimating the finite population mean. *Biom. J.*, 41, 5, 627-636.