Estimation of Population Variance in Simple Random Sampling using Auxiliary Information

Sumaira Ajmal Khan
Lahore Garrison University,
Phase 6, DHA Lahore

Mehwish Nawaz
Lahore Garrison University,
Phase 6, DHA Lahore

Kushbkht M. Din
Lahore Garrison University,
Phase 6, DHA Lahore

ABSTRACT
In this paper we propose a new estimator for the population variance using auxiliary information in simple random sampling . we derived a bias and mean square error equation of proposed estimator and compare with the bias and MSE of existing estimator and show that proposed estimator is more efficient than the existing estimators suggested by different authors such that Kadilar and Cingi (2005) [6] , Isaki(1983) [5]. We support this theoretical result with the help of a numerical illustration.

Keywords
Variance estimator, bias, MSE, simple random sampling, auxiliary information, Efficiency

1. INTRODUCTION
Auxiliary information has been used extensively in estimation of parameters like mean and variance since several decades.

Isaki (1983) [5] got inspiration from ratio estimator of finite population mean and proposed a ratio estimator , usually called classical estimator , of finite population variance. But Singh and Solanki (2013a) [12] claimed that Isaki (1983) [5] ratio estimator is the member of the class of estimators developed by Das and Tripathi (1978) [3]. Arcos and Rueda (1997) [2] suggested multivariate ratio estimator for population variance. Ahmed et al. (2000) [1] criticized the claim of Arcos and Rueda (1997) [2] , Kadilar and Cingi (2006a) [7] developed an estimator , ratio-type estimator of the mean of population, Kadilar and Cingi (2006b) [8] extended the idea of Isaki (1983) [5] ratio estimator for population variance . . Kadilar and Cingi (2006b) [8] involved the information available about coefficient of variation and coefficient of kurtosis of the auxiliary variable to generate these estimators under simple random sampling as well . Gupta and Shabbir (2008) [4] gave a hybrid class of variance estimators of population mean . Subramani and Kumarapandiyan (2012a) [9] modified the usual ratio-type estimator of Kadilar and Cingi (2006b) [8] for population variance using population median obtained from auxiliary variable . Subramani and Kumarapandiyan (2012b) [10] further modified the usual ratio-type variance estimators using lower and upper quartiles, inter-quartile range , quartile deviation and quartile average of the auxiliary variable . Subramani and Kumarapandiyan (2013) [11] developed another more efficient modified ratio-type estimator using median and coefficient of variation of the auxiliary variable.

2. NOTATIONS
N Population size n sample size

$\bar{Y} = \frac{\sum_{i=1}^{N} y_i}{N}$ population mean of the study variable y

$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}$ population mean of the auxiliary variable x

$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ sample mean of the study variable y

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ sample mean of the auxiliary variable x

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

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

$s_y^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{n-1}$ sample variance of the study variable

$s_x^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{n-1}$ sample variance of the auxiliary variable

$\lambda' = \frac{1}{n}$ $C_x, C_y =$ Coefficient of variations

$$\beta_{1(x)} = \frac{\mu_{03}^2}{\mu_{02}^2} \text{ Skewness of the auxiliary variable}$$

$$\beta_{2(x)} = \frac{\mu_{04}^2}{\mu_{02}^2} \text{ Kurttosis of the auxuliary variable}$$

$$\beta_{2(y)} = \frac{\mu_{40}^2}{\mu_{20}^2} \text{ Kurtosis of the study variable}$$

$$E(e_0) = 0 \quad , \quad E(e_1) = 0$$

3. EXISTING ESTIMATORS IN SIMPLE RANDOM SAMPLING

3.1 Isaki (1983) [5]
Motivated by the estimator of the population mean $\hat{\mu}_{yR} = \frac{\bar{y}}{\bar{x}}\bar{X}$ Isaki [5] suggested the ratio estimator of the variance of population using auxiliary information .

$$t_{Isaki} = s_y^2 \frac{S_x^2}{s_x^2} \quad \ldots\ldots\ldots\ldots.. \quad \textbf{(1)}$$

The MSE of the above estimator using first order approximation is

$\text{MSE}(t_{Isaki}) \cong \lambda' S_y^4 \{\beta_2(y) + \beta_2(x) - 2h\}$

Or

$\text{MSE}(t_{Isaki}) \cong \lambda' S_y^4 \{\beta_2'(y) + \beta_2'(x) - 2h'\}$

The bias is

$\text{Bias}(t_{Isaki}) \cong \lambda' S_y^2 \{\beta_2(x) - h\}$

Or

$\text{Bias}(t_{Isaki}) \cong \lambda' S_y^2 \{\beta_2'(x) - h'\}$

3.2 Kadilar and Cingi (2005) [6]
Kadilar and Cingi [6] proposed the following class of estimators for population variance .

$$t_{CH1} = s_y^2 [s_x^2 - \beta_2(x)]^{-1} [S_x^2 - \beta_2(x)]$$

$t_{CH2}= s_y^2[s_x^2 c_x - \beta_2(x)]^{-1}[S_x^2 c_x - \beta_2(x)]$
$t_{CH3}= s_y^2(s_x^2 - c_x)^{-1}(S_x^2 - c_x)$
$t_{CH4}= s_y^2[s_x^2\beta_2(x) - c_x]^{-1}[S_x^2\beta_2(x) - c_x]$

The MSEs of the above estimators are given below respectively :

MSE(t_{CH1}) $\cong$ $\lambda' S_y^4\left[\beta_2'(y) - 2\frac{S_x^2}{S_x^2-\beta_2(x)}h' + \left\{\frac{S_x^2}{S_x^2-\beta_2(x)}\right\}^2\beta_2'(x)\right]$

MSE(t_{CH2}) $\cong$ $\lambda' S_y^4\left[\beta_2'(y) - 2\frac{S_x^2}{S_x^2 C_x-\beta_2(x)}h' + \left\{\frac{S_x^2}{S_x^2 C_x-\beta_2(x)}\right\}^2\beta_2'(x)\right]$

MSE(t_{CH3}) $\cong$ $\lambda' S_y^4\left[\beta_2'(y) - 2\frac{S_x^2}{S_x^2-C_x}h' + \left\{\frac{S_x^2}{S_x^2-C_x}\right\}^2\beta_2'(x)\right]$

MSE(t_{CH4}) $\cong$ $\lambda' S_y^4\left[\beta_2'(y) - 2\frac{S_x^2}{S_x^2\beta_2(x)-C_x}h' + \left\{\frac{S_x^2}{S_x^2\beta_2(x)-C_x}\right\}^2\beta_2'(x)\right]$

The biases are

Bias(t_{CH1})= $\lambda' S_y^2\frac{S_x^2}{S_x^2-\beta_2(x)}\left[\frac{S_x^2}{S_x^2-\beta_2(x)}\beta_2'(x) - h'\right]$

Bias(t_{CH2})= $\lambda' S_y^2\frac{S_x^2}{S_x^2 C_x-\beta_2(x)}\left[\frac{S_x^2}{S_x^2-\beta_2(x)}\beta_2'(x) - h'\right]$

Bias(t_{CH3})= $\lambda' S_y^2\frac{S_x^2}{S_x^2-C_x}\left[\frac{S_x^2}{S_x^2-C_x}\beta_2'(x) - h'\right]$

Bias(t_{CH4})= $\lambda' S_y^2\frac{S_x^2}{S_x^2\beta_2(x)-C_x}\left[\frac{S_x^2}{S_x^2\beta_2(x)-C_x}\beta_2'(x) - h'\right]$

3.3 Subramani and Kumarapandiyan (2012b) [10]

Subramani and Kumarapandiyan [10] suggested the following efficient estimator

Using quartiles and functions of quartiles of the auxiliary variable.

$t_{jg1}=s_y^2(s_x^2 + Q_1)^{-1}(s_x^2 + Q_1)$
$t_{jg2}=s_y^2(s_x^2 + Q_3)^{-1}(s_x^2 + Q_3)$
$t_{jg3}=s_y^2(s_x^2 + Q_R)^{-1}(s_x^2 + Q_R)$
$t_{jg4}=s_y^2(s_x^2 + Q_D)^{-1}(s_x^2 + Q_D)$
$t_{jg5}=s_y^2(s_x^2 + Q_A)^{-1}(s_x^2 + Q_A)$

Where

$Q_R=Q_3$- Q_1

$Q_D = \frac{Q_3 - Q_1}{2}$ $\quad$ $Q_A=\frac{Q_3 + Q_1}{2}$

And their respective mean squared errors are:

MSE(t_{jg1}) =$\lambda' s_y^4\left[\beta_2'(y) - 2\frac{s_x^2}{s_x^2+Q_1}h' + \left\{\frac{s_x^2}{s_x^2+Q_1}\right\}^2\beta_2'(x)\right]$

MSE(t_{jg2}) =$\lambda' s_y^4\left[\beta_2'(y) - 2\frac{s_x^2}{s_x^2+Q_3}h' + \left\{\frac{s_x^2}{s_x^2+Q_3}\right\}^2\beta_2'(x)\right]$

MSE(t_{jg3}) =$\lambda' s_y^4\left[\beta_2'(y) - 2\frac{s_x^2}{s_x^2+Q_R}h' + \left\{\frac{s_x^2}{s_x^2+Q_R}\right\}^2\beta_2'(x)\right]$

MSE(t_{jg4}) =$\lambda' s_y^4\left[\beta_2'(y) - 2\frac{s_x^2}{s_x^2+Q_D}h' + \left\{\frac{s_x^2}{s_x^2+Q_D}\right\}^2\beta_2'(x)\right]$

MSE(t_{jg5}) =$\lambda' s_y^4\left[\beta_2'(y) - 2\frac{s_x^2}{s_x^2+Q_A}h' + \left\{\frac{s_x^2}{s_x^2+Q_A}\right\}^2\beta_2'(x)\right]$

And the respective biases are

B(t_{jg1}) =$\lambda' s_y^2\frac{s_x^2}{s_x^2+Q_1}\left[\frac{s_x^2}{s_x^2+Q_1}\beta_2'(x) - h'\right]$

B(t_{jg2}) =$\lambda' s_y^2\frac{s_x^2}{s_x^2+Q_3}\left[\frac{s_x^2}{s_x^2+Q_3}\beta_2'(x) - h'\right]$

B(t_{jg3}) =$\lambda' s_y^2\frac{s_x^2}{s_x^2+Q_R}\left[\frac{s_x^2}{s_x^2+Q_R}\beta_2'(x) - h'\right]$

B(t_{jg4}) =$\lambda' s_y^2\frac{s_x^2}{s_x^2+Q_D}\left[\frac{s_x^2}{s_x^2+Q_D}\beta_2'(x) - h'\right]$

B(t_{jg5}) =$\lambda' s_y^2\frac{s_x^2}{s_x^2+Q_A}\left[\frac{s_x^2}{s_x^2+Q_A}\beta_2'(x) - h'\right]$

4. PROPOSED ESTIMATOR

We proposed a new modified ratio type variance estimator of the auxiliary variable by using known value of kurtosis of the auxiliary variable . The modified ratio type variance estimator for population variance is defined as

$t_{new} = s_y^2\,\frac{S_x^2+S_x\beta_2(x)}{s_x^2+S_x\beta_2(x)}$

4.1 Bias of Proposed Estimator

$t_{new} = S_y^2(1+e_0)\frac{S_x^2+\boldsymbol{S_x\beta_2(x)}}{S_x^2(1+e_1)+\boldsymbol{S_x\beta_2(x)}}$

$$t_{new} = S_y^2(1+e_0)\left\{1+\frac{S_x^2 e_1}{S_x^2+S_x\beta_2(x)}\right\}^{-1}$$

$$t_{new} = S_y^2(1+e_0)\{1+\Omega e_1\}^{-1}$$

Where

$$\Omega = \frac{S_x^2}{S_x^2+S_x\beta_2(x)}$$

Expending the expression by using Taylor's series

$$t_{new} = S_y^2(1+e_0)\{1-\Omega e_1+(\Omega e_1)^2-(\Omega e_1)^3+\cdots\}$$

$$t_{new} = S_y^2(1+e_0)\{1-\Omega e_1+(\Omega e_1)^2\}$$

$t_{new} \quad -S_y^2 \quad = \quad S_y^2\{e_0-\Omega e_1+(\Omega e_1)^2-\Omega e_0 e_1\}$

E$\left(t_{new} \quad -S_y^2\right)$ = E$\left\{S_y^2(e_0-\Omega e_1+(\Omega e_1)^2-\Omega e_0 e_1)\right\}$

$$\text{B}(t_{new}) = \Omega\lambda' S_y^2\left\{\Omega\beta_{2(x)}'-h'\right\}$$

4.2 Mean Square Error of the proposed Estimator

Now comes MSE of the estimator

MSE(t_{new}) $\quad$ = E$\left\{S_y^2(e_0-\Omega e_1)\right\}^2$ $\qquad$ ignore higher order

$$= S_y^4\{E(e_0^2)+\Omega^2E(e_1^2)-2\Omega E(e_0e_1)\}$$

$$= S_y^4\left\{\lambda'\beta_{2(y)}'+\Omega^2\lambda'\beta_{2(x)}'-2\Omega\lambda' h'\right\}$$

$$=\lambda' S_y^4\left\{\beta_{2(y)}'+\beta_{2(x)}'\left(\Omega^2-2\Omega\frac{h'}{\beta_{2(x)}'}\right)\right\}$$

$$=\lambda' S_y^4\left\{\beta_{2(y)}'+\beta_{2(x)}'(\Omega^2-2\Omega v')\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(i)}$$

Where

$$\therefore \quad v' \quad = \quad \frac{h'}{\beta_{2(x)}'}$$

In order to minimize MSE to Differentiate (i) partially w.r.t Ω and equating to zero

$\frac{\partial}{\partial\Omega}$MSE($t_{new1}$) $\quad = \quad 0$

$\lambda' S_y^4\frac{\partial}{\partial\Omega}\left\{\beta_{2(y)}'+\beta_{2(x)}'(\Omega^2-2\Omega v')\right\} = 0$

$\Omega \quad = \quad v'$

$$= \lambda' S_y^4\{\beta'_{2(y)} - \beta'_{2(x)} v'^2\}$$

$$= \lambda' S_y^4 \beta'_2(y)\left\{1 - v'^2 \frac{\beta'_{2(x)}}{\beta'_{2(y)}}\right\}$$

$$= \lambda' S_y^4 \beta'_{2(y)}\left\{1 - \left(\frac{h'}{\sqrt{\beta'_{2(x)}\beta'_{2(y)}}}\right)^2\right\}$$

MSE(t_{new})= $\lambda' S_y^4 \beta'_{2(y)}\{1 - (\rho')^2\}$**……………..(b)**

Where

$$\rho' = \frac{h'}{\sqrt{\beta'_{2(x)}\beta'_{2(y)}}}$$

5. PROPOSED ESTIMATOR

To demonstrate the performance of the proposed estimator empirically in comparison to other estimators. I have used five data sets. The description of data sets is given below.

5.1 Data 1

Source: Gupta and Shabbir (2008) [4]
The available statistics are based on the data obtained from 104 villages existing in East Anatolia (Turkey) in 1999.
Y , the variable of interest represent apple 's production level per 100 tons
X , the auxiliary variable shows number of apple trees (in100s).

5.2 Data 2

Source : Gupta and Shabbir (2008) [4]
The data collected from 278 villages/Towns/wards under the control of Gajole polstation (Malda district , west Bengal) India.
Y , indicates number of agricultural labourers for the year 1971
X , the number of agricultural labourers for the year 1961

5.3 Data 3

Source : Rohini et al (Jan 2012)

Data collected from 142 Indian cities with population 0.1 million and above.
Y: 1971 population census
X: 1961 population census
n = 40

5.4 Data 4

Source: Kadilar and Cingi (2005) [6] ratio estimators for the population variance in simple random sampling.

The data obtained from 94 villages in Mediterranean(Turkey) in 1999.
Y: the variable of interest represents apple's production level per 100 tons
X: the auxiliary variable shows number of apple trees(in 100 sec).

5.5 Data 5

Source: Kadilar and Cingi (2005) [6] ratio estimators for the population variance in simple random sampling.

The available statistics obtained from 173 villages existing In East and southeast

Y: the variable of interest represents apple's production level per 100 tons

X: the auxiliary variable shows no of trees (1 unit shows 100 apple trees)

	Data 1	Data 2	Data 3	Data 4	Data 5
N	104	278	142	94	173
n	20	30	40	38	2
$\bar{Y}$	6.254	39.068	4015.2183	93.84	4.04
$\bar{X}$	13931.683	25.111	2900.3872	724.10	98.44
S_y	11.67	56.457167	8479.33801	299.07	9.46
S_x	23029.072	40.674797	6372.44072	1607.57	187.94
$\beta_2(y)$	16.523	25.8969	40.8536	24.14	27.96
$\beta_2(x)$	17.516	38.8898	48.1567	26.14	28.10
h	14.398	26.8142	43.7615	20.80	23.08
ρ	0.865	0.7213	0.9948	0.90	0.89

Table.1 Bias of the Existing and the Proposed Estimator

Estimators		Data 1	Data2	Data3	Data4	Dat5
Isaki (1983)	t_{Isaki}	21.23184951	1282.9966	7900281.139	12569.0763	224.623916
Kadilar & Cingi (2005)	t_{CH1}	21.23185393	1413.1212	7900391.028	12569.802	225.768988
	t_{CH2}	21.231852	1342.2774	7900326.9	12569.346	225.12942
	t_{TCH3}	21.23184993	21.231843	21.23184993	21.2318499	21.2318499
	t_{CH4}	21.231853	21.231853	21.231853	21.231853	21.231853
Subramani & Kumarapandiyan (2012b)	t_{jg1}	21.23184289	1074.8284	7900199.561	12568.417	222.857906
	t_{jg2}	21.23182966	737.70151	7900036.406	12567.0984	219.349828
	t_{jg3}	21.23183627	894.52749	7900117.983	12567.7577	221.099891
	t_{jg4}	21.23184289	1074.8284	7900199.561	12568.417	222.857906
	t_{jg5}	21.23183627	894.52749	7900117.983	12567.7577	221.099891
Proposed Estimator	t_{new}	21.13030158	- 350.02315	7210041.419	11436.3159	58.1996276

Table 2 MSE of the Existing and the Proposed Estimators

Estimators		Data 1	Data2	Data3	Data4	Data5
Isaki (1983)	t_{Isaki}	4862.205231	3778793.028	1.92215E+14	1827374379	39643.29503
Kadilar & Cingi (2005)	t_{CH1}	4862.205422	3983111.977	1.92216E+14	1827397122	39675.37353
	t_{CH2}	4862.885344	4862.885344	4862.885344	4862.885344	4862.885344
	t_{TCH3}	4862.205422	4862.205422	4862.205422	4862.205422	4862.205422
	t_{CH4}	4862.885467	4862.885467	4862.885467	4862.885467	4862.885467
Subramani & Kumarapandiyan (2012b)	t_{jg1}	4862.204945	3468925.171	1.92214E+14	1827353716	39594.00714
	t_{jg2}	4862.204372	3022246.855	1.92212E+14	1827312394	39496.7733
	t_{jg3}	4862.204659	3220379.202	1.92213E+14	1827333054	39545.16751
	t_{jg4}	4862.204945	3468925.171	1.92214E+14	1827353716	39594.00714
	t_{jg5}	4862.204659	3220379.202	1.92213E+14	1827333054	39545.16751
Proposed Estimator	t_{new}	4857.818785	2846652.986	1.84037E+14	1792753628	36249.92917

Table 3: PRE of the Existing and the Proposed Estimators

Estimators		Data 1	Data2	Data3	Data4	Data5
Isaki (1983)	t_{Isaki}	4862.205231	3778793.028	1.92215E+14	1827374379	39643.29503
Kadilar & Cingi (2005)	t_{CH1}	4862.205422	3983111.977	1.92216E+14	1827397122	39675.37353
	t_{CH2}	4862.885344	4862.885344	4862.885344	4862.885344	4862.885344
	t_{TCH3}	4862.205422	4862.205422	4862.205422	4862.205422	4862.205422
	t_{CH4}	4862.885467	4862.885467	4862.885467	4862.885467	4862.885467
Subramani & Kumarapandiyan (2012b)	t_{jg1}	4862.204945	3468925.171	1.92214E+14	1827353716	39594.00714
	t_{jg2}	4862.204372	3022246.855	1.92212E+14	1827312394	39496.7733
	t_{jg3}	4862.204659	3220379.202	1.92213E+14	1827333054	39545.16751
	t_{jg4}	4862.204945	3468925.171	1.92214E+14	1827353716	39594.00714
	t_{jg5}	4862.204659	3220379.202	1.92213E+14	1827333054	39545.16751
Proposed Estimator	t_{new}	4857.818785	2846652.986	1.84037E+14	1792753628	36249.92917

6. CONCLUSION

In this article we have proposed a modified ratio-type variance estimator using known value of kurtosis of the auxiliary variable. The bias and mean square error of the proposed modified ratio-type variance estimator are obtained and compared with that of existing modified ratio type variance estimator and show that proposed estimator is more efficient than the existing estimator. We have also assessed the performances of the proposed estimator for known population.

7. REFERENCES

[1] Ahmed, M.S. and Hossain, M.I. (2000) . Some competitive estimators of finite Population variance Multivariate Auxiliary Information, Information and Management Sciences, Volume 11 (1), 49-54

[2] Arcos , C.A. and Rueda , G.M , variance estimation using auxiliary information an almost unbiased multivariate ratio estimator. Matrika , Vol.45, pp.171-178 , 1997.

[3] Das, A.K, & Tripathi, T.P. (1978). Use of auxiliary information in estimating the finite population variance . Sankhya , 40, 139-148 .

[4] Gupta , S , & Shabbir, J. (2008). Variance estimation in simple random sampling using auxiliary information . Hacettepe Journal of Mathematics and Statistics, 37, 57-67.

[5] Isaki , C.T . (1983) . Variance estimation using auxiliary information . Journal of the American Statistical Association, 78, 117-123.

[6] Kadilar, C , Cingi ,H. (2005) . A new ratio estimator in stratified sampling . Comm. Statist. Theory Meth. 34:1-6.

[7] Kadilar , C , Cingi ,H. (2006a). Improvement in variance estimation using auxiliary information . Hacettepe Journal of Mathematics and Statistics , 35(1) , 111-115 .

[8] Kadilar , C , Cingi ,H. (2006b) . Ratio estimators for population variance in simple and stratified sampling . Applied Mathematics and Computation , 173, 1047-1058

[9] Subramani, J, & Kumarapandiyan , G. (2012a). Estimation of population mean using coefficient of variation and median of an auxiliary variable . International Journal of Probability and Statistics , 1(4) , 111-118 .

[10] Subramani, J, & Kumarapandiyan , G. (2012b) . Variance Estimation using median of an auxiliary variable. International Journal of Probability and Statistics, 1(3) , 36-40.

[11] Subramani, J, & Kumarapandiyan, G. (2013). Estimation of variance using known coefficient of variation and Median of an auxiliary variable. International Journal of Modern Applied Statistical Methods, 1(12) ,58-64.