Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 71(3) 2017 225–230

An Efficient Estimator for Estimating Population Variance using Auxiliary Information under Measurement Errors

Sheela Misra and Dipika
University of Lucknow, Lucknow

Received 12 June 2017; Accepted 27 June 2017

SUMMARY

In the present article an improved estimator for estimating finite population variance under measurement errors has been proposed. Bias and mean squared error of proposed estimator have been obtained up to first order of approximation. Theoretical efficiency comparison has been made among proposed estimator, sample variance estimator, usual ratio estimator and estimators proposed by Misra *et al.* (2016). A numerical illustration has been made using hypothetical data generated through R software. Interpretation of results is also shown through graphical representation of mean squared errors of estimators with and without measurement errors.

Keywords: Auxiliary information, Bias, Efficiency, Measurement errors, Mean squared error, Population variance, Simple random sampling, R Software.

1. INTRODUCTION

Generally in statistical analysis it is assumed that observations are collected without any error. However, in practice, this assumption may not be true and the data may be influenced by measurement errors due to various reasons, Cochran (1963), Sukhatme *et al.* (1984) and Biemer *et al.* (1991). When the observations are influenced by measurement errors then the estimates of population parameters (Mean, Variance, Total etc.) based on that values leads to the incorrect and misleading conclusions. So the study of consequences of measurement errors is essential. Measurement errors are generally taken as the discrepancy between true and observed values on any desirable characteristic. Measurement errors are generally taken as normally distributed with mean zero implies that average effect of measurement errors on respondents answer is zero, Biemer *et al.* (1991). But it will increase the variability, so study of effects of these errors needs attention. Many authors studied the effect of measurement errors on estimation of population parameters such as Maneesha and Singh (2001, 2002), Singh and Karpe (2008, 2009a, 2009b) and Diana and Giordan (2012). In the present article we are dealing about the estimation of finite population variance in the presence of measurement errors.

Let us consider Y and X as the study and auxiliary variables defined on a finite population $U\{= U_1, U_2, \ldots, U_N\}$ of size N and a sample of size n is taken by simple random sampling without replacement (SRSWOR) on these two characteristics Y and X. Here it is assumed that y_i and x_i are recorded instead of true values Y_i and X_i respectively. The observational errors /measurement errors are defined as

$$u_i = y_i - Y_i \tag{1.1.1}$$

$$v_i = x_i - X_i \tag{1.1.2}$$

u_i and v_i are random in nature with mean zero and different variances σ_u^2 and σ_v^2 respectively. It is assumed that u_i's and v_i's are uncorrelated although Y_i's and X_i's are correlated. Let (μ_Y, μ_X) and (σ_Y^2, σ_X^2) are mean and variances of (Y, X), i.e., study and auxiliary variables. ρ is the correlation coefficient between X and Y.

Corresponding author: Dipika
E-mail address: dipikascholar@gmail.com

Let $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,\ \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ be the unbiased estimators of the population means μ_Y and μ_X respectively.

$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ and $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ are not unbiased in presence of measurement errors.

The expected value of s_y^2 and s_x^2 under measurement errors are

$$E\left(s_y^2\right) = \sigma_Y^2 + \sigma_u^2 \text{ and } E\left(s_x^2\right) = \sigma_X^2 + \sigma_v^2$$

Let error variances σ_u^2 and σ_v^2 are known a prior than unbiased estimators of population variance under measurement errors are

$$\hat{\sigma}_y^2 = s_y^2 - \sigma_u^2 > 0$$

$$\hat{\sigma}_x^2 = s_x^2 - \sigma_v^2 > 0$$

Now, let us define

$$\hat{\sigma}_y^2 = \sigma_Y^2(1 + e_0)$$

$$\hat{\sigma}_x^2 = \sigma_X^2(1 + e_1)$$

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

$$E(e_0^2) = A_y / n,\ E(e_1^2) = A_x / n$$

$$E(e_0 e_1) = \delta - 1/n$$

Where,

$$A_y = \gamma_{2Y} + \gamma_{2U}\frac{\sigma_U^2}{\sigma_Y^2} + 2\left(1 + \frac{\sigma_U^2}{\sigma_Y^2}\right)^2$$

$$A_x = \gamma_{2X} + \gamma_{2V}\frac{\sigma_V^2}{\sigma_X^2} + 2\left(1 + \frac{\sigma_V^2}{\sigma_X^2}\right)^2$$

$$\delta = \frac{\mu_{22}(X, Y)}{\sigma_X^2 \sigma_Y^2}$$

$$\gamma_{2z} = \beta_{2z} - 3,\ \beta_{2z} = \frac{\mu_{4z}}{\mu_{2z}^2} \text{ and}$$

$$\mu_{rz} = E(z_i - \mu_z)^r;\ z = X, Y, U, V \text{ and}$$

$$r = 1, 2, 3, 4$$

$$\mu_{22}(XY) = E\{(X_i - \mu_X)^2 (Y_i - \mu_Y)^2\}$$

For estimating population variance the proposed estimator is

$$s_{y_\alpha}^2 = \hat{\sigma}_y^2\left[\frac{\hat{\sigma}_x^2}{\alpha\hat{\sigma}_x^2 + (1-\alpha)\,\sigma_x^2}\right] \tag{1.1.3}$$

Where α is the characterizing scalar chosen suitably.

2. BIAS AND MEAN SQUARE ERROR (MSE)

From (1.1.3), writing $s_{y_\alpha}^2$ in terms of e_i's

$$s_{y_\alpha}^2 = (1 + e_0)\sigma_y^2\left[\frac{(1 + e_1)\sigma_x^2}{\alpha(1 + e_1)\sigma_x^2 + (1 - \alpha)\sigma_x^2}\right]$$

$$(s_{ya}^2 - \sigma_y^2) = \sigma_y^2[e_0 + e_1 + e_0 e_1 - \alpha(e_1 + e_1^2 + e_0 e_1) + \alpha^2 e_1^2 + \ldots] \tag{1.2.1}$$

Taking expectation on both sides of (1.2.1), we get bias up to I[st] order of approximation

$$Bias(s_{y_\alpha}^2) = \sigma_y^2 E[e_0 + e_1 + e_0 e_1 - \alpha(e_1 + e_1^2 + e_0 e_1) + \alpha^2 e_1^2]$$

$$= \frac{\sigma_y^2}{n}(\alpha - 1)(1 - \delta - \alpha A_x) \tag{1.2.2}$$

Now, squaring (1.2.1) both sides and taking expectation, we have the mean square error of $s_{y_\alpha}^2$ up to first order of approximation to be

$$MSE(s_{y_\alpha}^2) = \sigma_y^4 E[e_0^2 + (\alpha - 1)^2 - 2(\alpha - 1)e_0 e_1]$$

$$MSE(s_{y_\alpha}^2) = \frac{\sigma_y^4}{n}[A_y + (\alpha - 1)^2 A_x - 2(\alpha - 1)(\delta - 1)] \tag{1.2.3}$$

The optimum value of α which minimizes the mean square error of $s_{y_\alpha}^2$ in (1.2.3) is given by

$$\alpha_0 = 1 - \frac{(\delta - 1)}{A_x} \tag{1.2.4}$$

The minimum value of mean square error of proposed estimator $s_{y_\alpha}^2$ for α_0 is given by

$$MSE(s_{y_\alpha}^2) = \frac{\sigma_y^4}{n}A_y - \frac{\sigma_y^4}{n}\frac{(\delta - 1)^2}{A_x} \tag{1.2.5}$$

Bias and MSE of existing estimators:

Table 1.1. Mean Squared Error of Existing estimator and proposed estimator

Estimators		Mathematical expression of estimator	Mean Square error with measurement errors
Sample Variance (t_0)		$\hat{\sigma}_y^2$	$\sigma_Y^2 \frac{A_x}{n}$
Isaki estimator (1983) (t_1)		$\hat{\sigma}_y^2\left(\frac{\sigma_X^2}{\hat{\sigma}_x^2}\right)$	$\frac{\sigma_Y^2}{n}(1+A_x-\delta)$
Kadilar and Cingi Estimator (t_2)		$\hat{\sigma}_y^2\left(\frac{\sigma_X^2-C_X}{\hat{\sigma}_x^2-C_X}\right)$	$\frac{\sigma_Y^4}{n}[A_y+B^2A_x-2B(\delta-1)]$
t_3	**Estimators Proposed by Misra *et al.* (2016)**	$\frac{\hat{\sigma}_y^2}{2}\left(1+\frac{\sigma_X^2}{\hat{\sigma}_x^2}\right)$	$\frac{\sigma_Y^4}{4n}[A_x+4\{A_y-(\delta-1)\}]$
t_4		$\hat{\sigma}_y^2\left(\frac{\sigma_X^2}{\hat{\sigma}_x^2}\right)^{1/2}$	
t_5		$\frac{2\hat{\sigma}_y^2}{1+\frac{\hat{\sigma}_x^2}{\sigma_X^2}}$	
t_6		$\frac{\hat{\sigma}_y^2}{2}\left[1+\frac{\sigma_X^2+C_X}{\hat{\sigma}_x^2-C_X}\right]$	$\frac{\sigma_Y^4}{4n}[4A_y+\mathrm{B}\{BA_x-4(\delta-1)\}]$
t_7		$\hat{\sigma}_y^2\left[\frac{\sigma_X^2+C_X}{\hat{\sigma}_x^2-C_X}\right]^{1/2}$	
t_8		$\frac{2\hat{\sigma}_y^2}{1+\frac{\hat{\sigma}_x^2+C_X}{\sigma_X^2-C_X}}$	
t_9		$\frac{\hat{\sigma}_y^2}{2}\left[\left(\frac{\sigma_X^2}{\hat{\sigma}_x^2}\right)+\frac{\sigma_X^2-C_X}{\hat{\sigma}_x^2-C_X}\right]$	$\frac{\sigma_Y^4}{4n}[4A_y+(1+\mathrm{B})\{4+(1+B)A_x-4\delta\}]$
t_{10}		$\hat{\sigma}_y^2\frac{\sigma_X}{\hat{\sigma}_x}\left(\frac{\sigma_X^2-C_X}{\hat{\sigma}_x^2-C_X}\right)^{1/2}$	
t_{11}		$\frac{2\hat{\sigma}_y^2}{\frac{\hat{\sigma}_x^2}{\sigma_X^2}+\frac{\hat{\sigma}_x^2-C_X}{\sigma_X^2-C_X}}$	
t_{12}		$\frac{\hat{\sigma}_y^2}{3}\left[1+\frac{\sigma_X^2}{\hat{\sigma}_x^2}+\frac{\sigma_X^2-C_X}{\hat{\sigma}_x^2-C_X}\right]$	$\frac{\sigma_Y^4}{4n}[9A_y-(1+\mathrm{B})\{6-(1+B)A_x-6\delta\}]$
t_{13}		$\hat{\sigma}_y^2\left[\left(\frac{\sigma_X^2}{\hat{\sigma}_x^2}\right)\left(\frac{\sigma_X^2-C_X}{\hat{\sigma}_x^2-C_X}\right)\right]^{1/3}$	
t_{14}		$\frac{3\hat{\sigma}_y^2}{1+\frac{\hat{\sigma}_x^2}{\sigma_X^2}+\frac{\hat{\sigma}_x^2-C_X}{\sigma_X^2-C_X}}$	
Proposed Estimator ($\hat{s}_{y_\alpha}^2$)		$\hat{\sigma}_y^2\left[\frac{\hat{\sigma}_x^2}{\alpha\hat{\sigma}_x^2+(1-\alpha)\sigma_x^2}\right]$	$\frac{\sigma_Y^4}{n}A_y-\frac{\sigma_Y^4}{n}\frac{(\delta-1)^2}{A_x}$

3. THEORETICAL EFFICIENCY COMPARISON OF PROPOSED ESTIMATOR WITH THE ESTIMATOR PROPOSED BY MISRA *et al.* (2016) EARLIER

(a) Efficiency comparison of proposed estimator $s^2_{y_\alpha}$ with sample variance t_0 is given by

$$MSE\left(s^2_{y_\alpha}\right) - MSE(t_0) < 0$$

$$\frac{\sigma_y^4}{n}A_y - \frac{\sigma_y^4}{n}\frac{(\delta-1)^2}{A_x} - \frac{\sigma_y^4}{n}A_y < 0$$

$$\mu_{22}(x,y) > \sigma_x^2\sigma_y^2 \tag{1.3.1}$$

(b) Efficiency comparison of proposed estimator $s^2_{y_\alpha}$ with Isaki estimator t_1 is given by

$$MSE\left(s^2_{y_\alpha}\right) - MSE(t_1) < 0$$

$$\frac{\sigma_y^4}{n}A_y - \frac{\sigma_y^4}{n}\frac{(\delta-1)^2}{A_x} - \frac{\sigma_y^4}{n}\left[A_y + A_x - 2(\delta-1)\right] < 0$$

$$(\delta - 1) < A_x \tag{1.3.2}$$

(c) Efficiency comparison of proposed estimator $s^2_{y_\alpha}$ with Isaki estimator t_2 is given by

$$MSE\left(s^2_{y_\alpha}\right) - MSE(t_2) < 0$$

$$\frac{\sigma_y^4}{n}A_y - \frac{\sigma_y^4}{n}\frac{(\delta-1)^2}{A_x} - \frac{\sigma_Y^4}{n}(2 + A_Y + A_X - 2\delta) < 0$$

$$(\delta - 1) < BA_x \tag{1.3.3}$$

(d) Efficiency comparison of proposed estimator $s^2_{y_\alpha}$ with Isaki estimator $MSE(t_3)/MSE(t_4)\,/MSE(t_5)$ is given by

$$MSE\left(s^2_{y_\alpha}\right) - MSE(t_3)/MSE(t_4)\,/MSE(t_5) < 0$$

$$\frac{\sigma_y^4}{n}A_y - \frac{\sigma_y^4}{n}\frac{(\delta-1)^2}{A_x} - \frac{\sigma_Y^4}{4n}\left[A_X + 4\{A_Y - (\delta-1)\}\right] < 0$$

$$2\delta - A_x < 2 \tag{1.3.4}$$

(e) Efficiency comparison of proposed estimator $s^2_{y_\alpha}$ with Isaki estimator $MSE(t_6)/MSE(t_7)\,/MSE(t_8)$ is given by

$$MSE\left(s^2_{y_\alpha}\right) - MSE(t_6)/MSE(t_7)\,/MSE(t_8) < 0$$

$$\frac{\sigma_y^4}{n}A_y - \frac{\sigma_y^4}{n}\frac{(\delta-1)^2}{A_x} - \frac{\sigma_Y^4}{4n}\left[4A_Y + B\{BA_X - 4(\delta-1)\}\right] < 0$$

$$2B(\delta - 1) > A_x \tag{1.3.5}$$

(f) Efficiency comparison of proposed estimator $s^2_{y_\alpha}$ with Isaki estimator $MSE(t_9)/MSE(t_{10})\,/MSE(t_{11})$ is given by

$$MSE\left(s^2_{y_\alpha}\right) - MSE(t_9)/MSE(t_{10})\,/MSE(t_{11}) < 0$$

$$\frac{\sigma_y^4}{n}A_y - \frac{\sigma_y^4}{n}\frac{(\delta-1)^2}{A_x} - \frac{\sigma_Y^4}{4n}\left[4A_Y + (1+B)\{4 + (1+B)A_X - 4\delta\}\right] < 0$$

$$2\delta - (1+B)A_x < 2 \tag{1.3.6}$$

(g) Efficiency comparison of proposed estimator $s^2_{y_\alpha}$ with Isaki estimator $MSE(t_{12})/MSE(t_{13})\,/MSE(t_{14})$ is given by

$$MSE\left(s^2_{y_\alpha}\right) - MSE(t_{12})/MSE(t_{13})\,/MSE(t_{14}) < 0$$

$$\frac{\sigma_y^4}{n}A_y - \frac{\sigma_y^4}{n}\frac{(\delta-1)^2}{A_x} - \frac{\sigma_Y^4}{9n}\left[9A_Y - (1+B)\{6 - (1+B)A_X - 6\delta\}\right] < 0$$

$$3(1-\delta)(1+B)^{-1} > A_x \tag{1.3.7}$$

Proposed estimator is better than usual unbiased estimator (sample variance), ratio estimator (Isaki estimator) and estimators proposed in presence of measurement errors if data satisfies the conditions (1.3.1)-(1.3.7).

4. NUMRICAL ILLUSTRATION

In this section, we demonstrate the performance of adopted estimators over other competitors, generating population from normal distribution by using R Software. The description of this data is as follows

$X = N(5,10)$, $Y = X + N(0,1)$, $y = Y + N(1,3)$, $x = X + N(1,3)$, $n = 5000$, $\mu_X = 4.95$, $\mu_Y = 4.93$, $\sigma_X^2 = 99.38$, $\sigma_Y^2 = 100.12$, $\sigma_u^2 = 25.57$, $\sigma_v^2 = 24.28$, $\rho_{XY} = 0.99$

The efficiencies of proposed estimator, usual unbiased estimator (sample variance), ratio estimator (Isaki) (1983), Kadilar and Cingi estimator (2006a) and estimators proposed by Misra *et al.* (2016), with and without measurement errors are:

Table: 1.2. MSE's of Estimators with and without measurement errors

Estimator	MSE with measurement errors	MSE without measurement errors
MSE (t_0)	6.25	3.93
MSE (t_1)	4.60	0.97
MSE (t_2)	4.68	0.09
MSE (t_3)/MSE (t_4)/MSE (t_5)	3.90	1.02
MSE (t_6)/MSE (t_7)/MSE (t_8)	3.88	0.98
MSE (t_9)/MSE (t_{10})/MSE (t_{11})	4.78	2.02
MSE (t_{12})/MSE (t_{13})/MSE (t_{14})	3.85	0.473
MSE($s^2_{y\alpha}$)	3.78	0.081

The graphical Comparison of Estimators: The graphical representation of MSE's of Proposed and existing estimators are shown in Fig. 1.1.

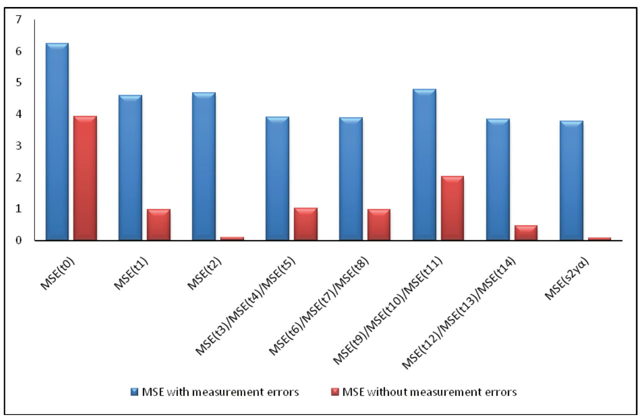

Fig. 1.1 Bar graph of MSE's of Proposed Estimator and existing estimators

5. CONCLUDING REMARKS

(1) The minimum value of mean square error for the optimum value of α clear in (1.2.5) is

$$MSE\left(s^2_{y_\alpha}\right) = \frac{\sigma_y^4}{n}A_y - \frac{\sigma_y^4}{n}\frac{(\delta-1)^2}{A_x}$$

This always results in the reduction of mean square error (variance) of sample variance.

(2) Conditions are derived in (1.3.1) to (1.3.7) under which proposed estimator performed better than estimator defined earlier.

(3) From Table (1.2), we can conclude that the performance of proposed estimator is better than all other existing estimators which are considered under measurement errors i.e. if measurement errors are present in study as well as in auxiliary variable than using proposed estimator provides better results than other estimators.

(4) Also the conclusion 3 is supported by graphically (Fig. 1.1) i.e. the proposed estimator is better than other estimators with and without measurement errors both.

REFERENCES

Biemer, Paul P., Groves, Robert M. Lars E. Lyberg, Lars E., Mathiowetz, Nancy A. and Sudman, Seymour (1991). *Measurement Errors in Surveys*, Wiley, New York.

Cochran, W.G. (1963). *Sampling Techniques*. Second Edition, Wiley Eastern Private Limited, New Delhi.

Das, A.K. and Tripathi, T.P. (1978). Use of auxiliary information in estimating the finite population variance, *Sankhya*, **c4**,139-148.

Diana, G. and Giordan, M. (2012). Finite population variance estimation in presence of measurement errors. *Comm. Statist. Theory. Methods*, **41**, 4302-4314.

Gujrati, D.N. and Sangeeta (2007). *Basic Econometrics*. McGraw-Hill.

Isaki, C.T. (1983). Variance estimation using Auxiliary Information. *J. Amer. Statist. Asssoc*., **78**, 117-123.

Kadilar, C. and Cingi, H. (2006a). Improvement in variance estimation using auxiliary information. *Hacettepe J. Maths. Statist*., **35(1)**, 111-115.

Maneesha and Singh, R.K. (2001). An estimation of population mean in the presence of measurement errors. *J. Ind. Soc. Agril. Statist.*, Vol. 54.

Maneesha and Singh, R.K. (2002). Role of regression estimator involving measurement errors. *Brazilian J. Probab. Statist.*, **16**, 39-46.

Misra, *et al.* (2016). Some improved estimators for estimating population variance in the presence of measurement errors. *J. Stat. Appl. Pro*., **5(2)** (Jul. 2016), 311-320.

Singh, H.P., Singh, A.K. and Solanki, R.S. (2014), Estimation of finite population variance using auxiliary information in sample surveys. *Statistica*, **XXIV(1)**, 99-116.

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C. (1984). *Sampling Theory of Surveys with Applications*. Iowa State University Press, AMES.

Shalabh (1997). Ratio method of estimation in the presence of measurement errors. *J. Ind. Soc. Agril. Statist.*, **50(2)**, 150-155.

Sharma, P. and Singh, R. (2013). A generalized class of estimators for finite population variance in presence of measurement error. *J. Morden Appl. Statist. Methods*, **12(2)**, 231-241.

Singh, H.P. and Karpe, N. (2008). Estimation of population variance using auxiliary information in the presence of measurement errors. *Statist. Trans.*- new series, **9(3)**, 443-470.

Singh, H. and Karpe, N. (2009a). A class of estimators using auxiliary information for estimating finite population variance in presence of measurement errors. *Comm. Statist. – Theory Methods*. **38**, 734–741.

Singh, H. and Karpe, N. (2009b). A general procedure for estimating the general parameter using auxiliary information in presence of measurement errors. *Comm. Korean Statist. Soc*. **16(5)**, 821-840.

Srivastava, S.K. and Jhajj, H.S. (1980). A class of estimators using auxiliary information for estimating finite population variance. *Sankhya*, **C4**,87-96.