International Journal of Mathematics And its Applications
Volume 4, Issue 2–D (2016), 23–28.
ISSN: 2347-1557
Available Online: http://ijmaa.in/

International Journal of Mathematics And its Applications

An Efficient Estimator for Estimating Population Variance in Presence of Measurement Errors

Research Article

Sheela Misra[1], Dharmendra Kumar Yadav[1]* and Dipika[1]

1 Department of Statistics, University of Lucknow, Lucknow, India.

Abstract: The present paper advocates the problem of estimating population variance when the measurement errors are present in both the study variable and the auxiliary variable.Bias and Mean Square Error (MSE) of the proposed estimator is obtained up to first order of approximation. Theoretical efficiency comparison between usual variance estimator and the proposed estimator is also made under measurement errors.Theoretical results are supported by simulation study using R software.

Keywords: Variance, Measurement Errors, Bias, MSE, Efficiency, Simulation, R software.

1. Introduction

Most of the statistical procedures for the data analysis assume that the observations on the characteristic of interest are recorded without any error. But in real life such assumptions are not valid and data is contaminated with measurement errors due to various reasons [1, 2, 4]. When observations are subject to measurement errors then the estimates of population parameters (Mean, Variance, Total etc.) based on that values leads to the incorrect estimates. So the study of these errors is much needed.

Measurement errors play a very significant role especially in surveys which are related with socially undesirable characteristics and may mislead in conclusions if not handled properly. Measurement errors arise in data collection or taking observations and are mainly contributed by the respondent or the enumerator or both. Measurement errors refer to the discrepancy between the individual true value and the corresponding observed sampling value irrespective of reasons for discrepancies. Since these errors arise at the time of observation taking process so sometimes these are also known as observational errors. Responses of interviewers are also one of the reasons of these errors, so these are also known as Response errors. Measurement errors or observational errors are sometimes called as accommodation errors.

Measurement errors are generally taken as normally distributed with mean zero implies that average effect of measurement errors on respondents answer is zero, chapter 1 by M. Groves ([6]). But it will increase the variability, so estimation of these errors needs a wide attention. Many authors including Das [17], Srivastava [15], Singh [8] and Diana [18], studied the effects of measurement errors on estimation of population parameters. In the present article we are focusing upon the study of estimation of population variance when the observations are subject to measurement errors.

* *E-mail: dkumar.yadava@gmail.com*

2. Estimation of Population Variance Under Measurement Errors

Let us assume that Y and X are 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 drawn 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 observed instead of true values Y_i and X_i respectively. The measurement errors are defined as

$$u_i = y_i - Y_i \tag{1}$$

$$v_i = x_i - X_i \tag{2}$$

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. It is also assumed that $u_i's$ and $v_i's$ are uncorrelated with $Y_i's$ and $X_i's$ respectively. Let (μ_Y, μ_X) and $(\sigma_Y^2,\ \sigma_X^2)$ are mean and variances of $(Y,\ X)$, i.e, study and auxiliary variables. ρ is the correlation coefficient between X and Y. Let $\bar{y} = \frac{1}{n}\sum\limits_{i=1}^{n} y_i$, $\bar{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$ be the unbiased estimators of the population means μ_Y and μ_X respectively.

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

and

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

are not unbiased estimators of the population variances σ_X^2 and σ_Y^2. In presence of measurement errors the expected value of s_y^2 and s_x^2 is given by $E(s_y^2) = \sigma_Y^2 + \sigma_u^2$ and $E(s_x^2) = \sigma_X^2 + \sigma_v^2$. 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

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

$$\bar{x} = \mu_X(1 + e_1)$$

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

$$E(e_0^2) = \frac{A_y}{n}$$

$$E(e_1^2) = \frac{C_X^2}{n\theta_X}$$

$$E(e_0e_1) = \frac{\delta C_X}{n}$$

Where,

$$A_y = \gamma_{2Y} + \gamma_{2u}\frac{\sigma_u^2}{\sigma_Y^2} + 2(1 + \frac{\sigma_u^2}{\sigma_Y^2})^2$$

$$\theta_x = \frac{\sigma_X^2}{\sigma_X^2 - \sigma_v^2}$$

$$\delta = \frac{\mu_{12}(X,Y)}{\sigma_X \sigma_Y^2}$$

$$\gamma_{2z} = \beta_{2z} - 3,$$

$$\beta_{2z} = \mu_{4z}/(\mu_{2z}^2) \text{ and}$$
$$\mu_{rz} = E(z_i - \mu_z)^2;$$
$$z = X, Y, U, V$$
$$\mu_{22}(XY) = E\{(X_i - \mu_X)^2 (Y_i - \mu_Y)^2\}$$

The usual unbiased estimator of the population variance of the study variable Y under measurement errors is defined by

$$t_0 = \hat{\sigma}_y^2 \tag{3}$$
$$= \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$
$$Bias(t_0) = 0 \tag{4}$$
$$MSE(t_0) = \sigma_Y^4 \frac{A_y}{n} \tag{5}$$

3. Suggested Estimator

$$t_k = \hat{\sigma}_y^2[1 + k\frac{(\bar{x} - \mu_X)}{\mu_X}] \tag{6}$$
$$= \sigma_y^2(1 + e_0)[1 + k\frac{(\mu_x(1 + e_0) - \mu_x)}{\mu_x}]$$
$$= \sigma_y^2(1 + ke_1 + e_0 + ke_0e_1)$$
$$(t_k - \sigma_y^2) = \sigma_y^2(ke_1 + e_0 + ke_0e_1) \tag{7}$$
$$Bias(t_k) = \sigma_y^2 kE(e_0e_1)$$
$$Bias(t_k) = \sigma_y^2 k\frac{\delta C_X}{n} \tag{8}$$

hence

$$Bias(t_k) = \frac{\mu_{12}(X, Y)}{n\mu_x}$$

Which is of order $O(1/n)$, hence the bias of estimator t_k is negligible for sufficient large value of n, and for symmetrical population bias becomes zero for the order of our approximation. Squaring Equation (7) both sides, we get

$$(t_k - \sigma_y^2)^2 = \sigma_y^4[ke_1 + e_0 + ke_0e_1]^2$$
$$= \sigma_y^4[k^2e_1^2 + e_0^2 + k^2e_0^2e_1^2 + 2ke_0e_1 + 2k^2e_0e_1^2 + 2ke_1e_0^2]$$

Taking expectation on both sides and ignoring terms of higher order we get the mean square error (MSE)

$$= E(t_k - \sigma_y^2)^2 = \sigma_y^4 E[k^2e_1^2 + e_0^2 + 2ke_0e_1]$$
$$E(t_k - \sigma_y^2)^2 = \sigma_y^4[k^2E(e_1^2) + E(e_0^2) + 2kE(e_0e_1)]$$

Substituting the values of $E(e_0^2)$, $E(e_1^2)$ and $E(e_0e_1)$ we get

$$MSE(t_k) = \frac{\sigma_y^4}{n}\left[k^2\frac{C_X^2}{\theta_X} + A_Y + 2k\lambda C_x\right]$$

$$MSE(t_k) = \frac{\sigma_y^4}{n}\left[k^2 C_X^2 (1+\frac{\sigma_v^2}{\sigma_X^2}) + A_Y + 2k\lambda C_x\right] \tag{9}$$

Now the optimum value of k is

$$k = -\lambda \frac{\theta_X}{C_X} \tag{10}$$

Hence minimum MSE is obtained by substituting the value of k in (9)

$$\begin{aligned} MSE(t_k)_{\min} &= \frac{\sigma_Y^4}{n}(2+\gamma_Y^2) + \frac{\sigma_Y^4}{n}\frac{\sigma_u^2}{\sigma_Y^4}(2+\gamma_u^2) + 4\frac{\sigma_u^2}{\sigma_Y^2} - \lambda^2\left(\frac{\sigma_X^2}{\sigma_v^2+\sigma_X^2}\right) \\ &= \text{Sampling variance } (V_1) + \text{Measurement variance } (V_2) \end{aligned} \tag{11}$$

where $V_1 = \frac{\sigma_Y^4}{n}(2+\gamma_Y^2)$ and $V_2 = \frac{\sigma_Y^4}{n}\left[\frac{\sigma_u^2}{\sigma_Y^4}(2+\gamma_u^2) + 4\frac{\sigma_u^2}{\sigma_Y^2} - \lambda^2\left(\frac{\sigma_X^2}{\sigma_v^2+\sigma_X^2}\right)\right]$.

4. Efficiency Comparison with Usual Unbiased Estimator

Proposed estimator will be more efficient than the usual unbiased estimator iff $MSE(t_k)_{\min} - MSE(t_0) < 0$ which provides

$$\lambda^2\left(\frac{\sigma_X^2}{\sigma_v^2+\sigma_X^2}\right) > 0 \tag{12}$$

Our proposed estimator t_k will be more efficient than the usual unbiased variance estimator t_0 if the condition (12) is satisfied by the data set.

5. Simulation Study

In this section, we demonstrate the performance of adopted estimator over the usual variance estimator in presence of measurement errors, generating population from normal distribution by using R programme. The description of this data is as follows $X = N(5,10)$, $Y = X + N(0,1)$, $y = Y + N(1,3)$, $x = X + N(1,3)$, $n = 5000$, $\mu_X = 4.95$, $\mu_Y = 4.93$, $\sigma_X^2 = 99.38$, $\sigma_Y^2 = 100.12$, $\sigma_u^2 = 25.57$, $\sigma_v^2 = 24.28$, $\rho_{XY} = 0.99$. By using these values we find that the condition (12) is satisfied by the above data set. So our proposed estimator will be more efficient than the usual unbiased estimator. Now the MSE's of both estimators are given as

$$MSE(t_k)_{\min} = 6.01$$
$$MSE(t_0) = 6.25$$

6. Concluding Remarks

(a). Since the expression of Bias does not involve σ_u^2 or σ_v^2 so Bias of the proposed estimator is not affected by the measurement errors.

(b). Since $MSE(t_k)_{\min} < MSE(t_0)$ so our proposed estimator is more efficient than the usual unbiased estimator in presence of measurement errors.

References

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

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

[3] T.Lessler Judith, Kalsbeek and D.William, *Non-Sampling Error in Surveys*, John Wiley and Sons. Inc, (2012).

[4] P.V.Sukhatme, B.V.Sukhatme, S.Sukhatme and C.Ashok, *Sampling Theory of Surveys with Applications*, Iowa State University Press, Ams, USA and Indian Society of Agricultural Statistics, New Delhi, (1984).

[5] Morris H.Hansen, William N.Hurwitz, Eli S.Marks and W.Parker Mauldin, *Response Errors in Survey*, Journal of American Statistical Association, 46(254)(1951), 147-190.

[6] Paul P.Biemer, Robert M.Groves, Lars E.Lyberg, Nancy A.Mathiowetz and Seymour Sudman, *Measurement Errors in Surveys*, John Wiley & Sons, (2004).

[7] P.C.Mahalanobis, *Recent Experiments in Statistical Sampling in the Indian Statistical Institute*, Journal of Royal Statistical Society, 109(1946), 327-328.

[8] H.P.Singh and N.Karpe, *A General Procedure for Estimating the General Parameter Using Auxiliary Information in Presence of Measurement Errors*, Communication of the Korean Statistical Society, 16(5)(2009), 821-840.

[9] Salabh, *Ratio Method of Estimation in the Presence of Measurement Errors*, Journal of the Indian Society of Agricultural Statistics, 1(2)(1997), 150-155.

[10] Rachna Maithani, *On Some Classes of Estimators in Sampling Theory Using Auxiliary Information*, Ph.D thesis, submitted and awarded to the Department of Statistics, University of Lucknow, (2013).

[11] P.Sharma and R.Singh, *A Generalized class of estimators for finite population variance in presence of measurement error*, Jour. Morden Applied Statistical Methods, 12(2)(2013), 231-241.

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

[13] H.P.Singh and N.Karpe, *A class of estimators using auxiliary information for estimating finite population variance in presence of measurement errors*, Commun. Statist. Theo. Meth., 38(2009), 734741.

[14] P.Maiti, *Estimation of Non Sampling Variance Components Under the Linear Model Approach*, Statistics in Transition-new series, 10(2)(2009), 193-222.

[15] S.K.Srivastava and H.S.Jhajj, *A class of estimators using auxiliary information for estimating finite population variance*, Sankhya, c(4)(1980), 87-96.

[16] P.V.Sukhatme and G.R.Seth, *Non Sampling Errors in Surveys*, Journal of the Indian Society of Agricultural Statistics, 4(1952), 5-41.

[17] A.K.Das and T.P.Tripathi, *Use of auxiliary information in estimating the finite population variance*, Sankhya, c(4)(1978), 139-148.

[18] G.Diana and M.Giordan, *Finite population variance estimation in presence of measurement errors*, Communication in Statistics Theory and Methods, 41(2012), 4302-4314.

[19] C.T.Isaki, *Variance estimation using Auxiliary Information*, Journal of American Statistical Association, 78(1983), 117-123.

[20] C.Kadilar and H.Cingi, *Improvement in Variance estimation using Auxiliary Information*, Hacettepe Journal of Mathematics and Statistics, 35(1)(2006), 111-115.

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

[22] S.Misra and D.K.Yadav, *Estimating Population Mean Using Known Coefficient of Variation under Measurement Errors in the edited book Statistics and Informatics in Agricultural Research*, Excel India Publisher, New Delhi, (2016).

[23] S.Misra, Dipika and D.K.Yadav, *Some Improved Estimators for Estimating Population Variance in the Presence of Measurement Errors*, Journal of Statistics Applications and Probability, 5(2)(2016), 311-320.