Special Issue of Engineering and Scientific International Journal ISSN 2394-7187(Online) ISSN 2394-7179 (Print)
*Proceedings of the UGC Sponsored - National Conference on "**Recent Statistical Computing Techniques and their Applications**"*
Department of Statistics and Department of Computer Science, Ramanujan College,University of Delhi, Delhi.
March 11-12, 2016. ***Published on January 2017***

An Improved Estimator for Estimating Population Mean in Presence of Measurement Errors

Sheela Misra[#1], Dharmendra Kumar Yadav[*2], Dipika[*3]

[1]*Department of Statistics, University of Lucknow, Lucknow, 226007, India*
profsheelamisra@gmail.com
[2]*Department of Statistics, University of Lucknow, Lucknow,226007, India*
dkumar.yadava@gmail.com
[3]*Department of Statistics, University of Lucknow, Lucknow, 226007, India*
dipikascholar@gmail.com

Abstract— The present manuscript examines the effects of measurement errors on a regression type estimator used for estimation of finite population mean. The bias and mean square error (MSE) of the proposed estimator are obtained up to first order of approximation. Theoretical Efficiency comparison is also done between the proposed estimator and the usual linear regression estimator. The results have been illustrated by carrying out the simulation study using R software.

Keywords— *Measurement Errors, Regression Estimator, Bias, Mean Square Error, Efficiency, Simulation, R Software*

1. Introduction

In most of the statistical procedures the analysis of data is based on the assumption that the observations on the characteristics to be studied are recorded without any error. However, in practical situations this assumption is not fulfilled by the data set. Observations, collected on some characteristic are quite different from their true values. Such discrepancy is referred as measurement errors or observational errors. In sample surveys, such errors are very common and affect the estimation procedures adversely in terms of increased bias and variability. So the study of such consequences of measurement errors is essential. Several authors including Shalabh (1997), Sud and Srivastava (2000), Maneesha and Singh (2001, 2002), Srivastava and Shalabh (2001), Allen et al.(2003), Singh and Karpe (2008a, 2008b, 2009), Maiti (2009) , Kumar et al (2011) , Shukla et al (2012) etc studied the effects of measurement errors on estimation of population parameters . In the present article we study the estimation of population mean in presence of measurement errors.

2. Notations

Suppose that we are given a finite population U = {U_1, U_2,…….U_N} of size N. Further assume that Y and X are the study and the auxiliary variables respectively. A set of n paired observations is obtained from the given population through simple random sampling without replacement (SRSWOR) procedure on two characteristics X and Y. It is assumed that x_i and y_i for the i^{th} sampling unit are measured instead of true values X_i and Y_i . The observational or measurement errors are defined as

$$u_i = y_i - Y_i$$
$$v_i = x_i - X_i$$

which are assumed to be stochastic with mean zero and different variances σ_u^2 and σ_v^2. We assume that although X_i 's and Y_i's are correlated, but the correlation between u_i's and v_i's is zero. It is also assumed that measurement errors are uncorrelated with true values of x and y.

Let (μ_X, μ_Y) and (σ_X^2, σ_Y^2) be the population means and population variances of the characteristics X and Y respectively. Let ρ be the population correlation coefficient between X and Y. Let $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ be the unbiased estimators of population means μ_X and μ_Y respectively.

We note that $s_x^2 = \frac{1}{n-1}\sum_{i=n}^{n}(x_i - \bar{x})^2$ and $s_y^2 = \frac{1}{n-1}\sum_{i=n}^{n}(y_i - \bar{y})^2$ 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 $E(s_y^2) = \sigma_Y^2 + \sigma_u^2$, $E(s_x^2) = \sigma_x^2 + \sigma_v^2$.

Let error variances σ_u^2 and σ_v^2 are known a prior then unbiased estimators of population variance in presence of measurement errors are

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

We further assume the following approximations

$$\bar{y} = \mu_Y(1 + e_0), \bar{x} = \mu_X(1 + e_1)$$
$$\hat{\sigma}_y^2 = \sigma_y^2(1 + e_2), \hat{\sigma}_x^2 = \sigma_x^2(1 + e_3)$$
$$\hat{\sigma}_{xy} = \sigma_{xy}(1 + e_4)$$

such that $E(e_0) = E(e_1) = E(e_2) = E(e_3) = E(e_4) = 0$

From Singh and Karpe (2009) ,we have ,

$E(e_0^2) = \frac{C_Y^2}{n\theta_Y}$, $C_Y = \frac{\sigma_Y}{\mu_Y}$, $C_X = \frac{\sigma_X}{\mu_X}$, $\theta_X = \frac{\sigma_X^2}{\sigma_X^2+\sigma_Y^2}$, $\theta_Y = \frac{\sigma_Y^2}{\sigma_Y^2+\sigma_u^2}$

$E(e_1^2) = \frac{C_X^2}{n\theta_X}$, $E(e_2^2) = \frac{A_Y}{n}$,

$E(e_3^2) = \frac{A_X}{n}$

where , $A_Y = \gamma_{2Y} + \gamma_{2u}\frac{\sigma_u^4}{\sigma_Y^4} + 2(1+\frac{\sigma_u^2}{\sigma_Y^2})^2$

$A_X = \gamma_{2X} + \gamma_{2v}\frac{\sigma_v^4}{\sigma_X^4} + 2(1+\frac{\sigma_v^2}{\sigma_X^2})^2$

$E(e_0e_1) = \rho\frac{C_XC_Y}{n}$, $E(e_1e_2) = \frac{\mu_{1200}}{n\sigma_X^2\mu_X}$

$E(e_1e_3) = \frac{\mu_{3000}}{n\sigma_X^2\mu_X}$, $E(e_0e_2) = \frac{\mu_{0300}}{n\sigma_Y^2\mu_Y}$

$E(e_1e_4) = \frac{\mu_{2100}}{n\sigma_{XY}\mu_X}$

$\mu_{pqrs} = E(X-\mu_X)^p(Y-\mu_Y)^q u^r v^s$

3. Estimation of Finite Population Mean under measurement errors

We are studying the performance of following estimator used for estimation of finite population mean in presence of measurement errors.

$$\bar{y}_k = \bar{y} + b(\mu_X - \bar{x}) + k\{\bar{y} - (\hat{\sigma}_y/C_Y)\} \quad (3.1)$$

where k is characterizing scalar to be chosen suitably and , $b = \hat{\sigma}_{xy}/\hat{\sigma}_x^2$ = Regression coefficient.

By using notations , $b = \hat{\sigma}_{xy}/\hat{\sigma}_x^2$ can be written as

$$b = \frac{\sigma_{XY}(1+e_4)}{\sigma_X^2(1+e_3)}$$

$$= \frac{\sigma_{XY}}{\sigma_X^2}(1+e_4)(1-e_3+e_3^2-\cdots\ldots)$$

$$= \frac{\sigma_{XY}}{\sigma_X^2}(1-e_3+e_3^2+e_4-e_3e_4)$$

Now the proposed estimator can be written as

$$\bar{y}_k = \mu_Y(1+e_0) + \frac{\sigma_{XY}(1+e_4-e_3-e_3e_4+e_3^2)\{\mu_X-\mu_X(1+e_1)\}}{\sigma_X^2} + k[\mu_Y(1+e_0) - \frac{(s_y^2)^{\frac{1}{2}}}{C_Y}]$$

$$= \mu_Y + e_0\mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}(1+e_4-e_3-e_3e_4 + e_3^2)(-\mu_Xe_1) + \frac{k}{C_Y}\left[\mu_Y(1+e_0)C_Y - (1+e_2)^{\frac{1}{2}}\sigma_Y\right]$$

$$\bar{y}_k - \mu_Y = e_0\mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}\mu_X[-e_1 - e_1e_4 + e_3e_1 + e1e3e4 - e1e32 + k\mu Y[e0 - 12e2 + 18e22 \quad (3.2)$$

taking expectation on both sides in (3.2) and using first order approximation

$$E(\bar{y}_k - \mu_Y) = \frac{\sigma_{XY}}{\sigma_X^2}\mu_X\{E(e_1e_3) - E(e_1e_4)\} + k\mu_Y\left\{\frac{1}{8}E(e_2^2)\right\}$$

Substituting the values of $E(e_1e_3)$, $E(e_1e_4)$ and $E(e_2^2)$, we get the Bias of $\bar{y}_k$ upto the terms of $O(\frac{1}{n})$,

$$\text{Bias}(\bar{y}_k) = \frac{1}{n}\left[\frac{k\mu_Y A_Y}{8} - \frac{\sigma_{XY}}{\sigma_X^2}\left\{\frac{\mu_{2100}}{\sigma_{XY}} - \frac{\mu_{3000}}{\sigma_X^2}\right\}\right] \quad (3.3)$$

squaring and taking expectation of (3.2) on both sides , we get the MSE of proposed estimator $\bar{y}_k$ up to the order $O\left(\frac{1}{n}\right)$ to be

$$E(\bar{y}_k - \mu_Y)^2 = \mu_Y^2E(e_0^2) + \left(\frac{\sigma_{XY}}{\sigma_X^2}\right)^2\mu_X^2E(e_1^2) - 2\left(\frac{\sigma_{XY}}{\sigma_X^2}\right)\mu_X\mu_YE(e_0e_1) + k^2\mu_Y^2\left[E(e_0^2) + \frac{1}{4}E(e_2^2) - E(e0e2) + 2k\mu Y2Ee02 - 12E(e0e2) - 2k\sigma XY\sigma X2\mu X\mu Y[Ee0e1 - 12E(e1e2)]\right]$$

Substituting the values of $E(e_0^2), E(e_1^2), E(e_2^2), E(e_0e_1), E(e_0e_2), E(e_1e_2)]$ we get MSE

$$\text{MSE}(\bar{y}_k) = \mu_Y^2\left(\frac{C_Y^2}{n\theta_Y}\right) + \left(\frac{\sigma_{XY}}{\sigma_X^2}\right)^2\mu_X^2\left(\frac{C_X^2}{n\theta_X}\right) - 2\left(\frac{\sigma_{XY}}{\sigma_X^2}\right)\mu_X\mu_Y\left(\frac{\rho C_XC_Y}{n}\right) + k^2\mu_Y^2\left[\left(\frac{C_Y^2}{n\theta_Y}\right) + \frac{1}{4}\left(\frac{A_Y}{n}\right) - \frac{\mu_{0300}}{n\sigma_Y^2\mu_Y}\right] + 2k\mu_Y^2\left[\left(\frac{C_Y^2}{n\theta_Y}\right) - 12\mu0300n\sigma Y2\mu X - 2k\sigma XY\sigma X2\mu X\mu Y\rho CXCYn - 12\mu1200n\sigma Y2\mu X$$

On solving we get

$$\text{MSE}(\bar{y}_k) = \frac{\sigma_Y^2}{n}(1-\rho^2) + \frac{k^2\mu_Y^2}{4n}\left[4C_Y^2 + A_Y - 4\mu0300\sigma Y2\mu Y + kn \quad 2\sigma u2 + 2\sigma Y2 - \mu0300\sigma YCY - 2\rho2\sigma X\sigma Y + \rho\mu1200CY\sigma Y + 1n\sigma u2 + \rho2\sigma v2\sigma X2\sigma Y2\right] \quad (3.4)$$

Now optimizing MSE $(\bar{y}_k)$, we get the optimum value of k

$$k_{opt} = \frac{-2\left[\frac{\rho\mu_{1200}}{C_Y\sigma_Y} + 2(\sigma_u^2+\sigma_Y^2) - \frac{\mu_{0300}}{\sigma_YC_Y} - 2\rho^2\sigma_X\sigma_Y\right]}{\mu_Y^2\left[\frac{4C_Y^2}{\theta_Y} + A_Y - 4\frac{\mu_{0300}}{\sigma_Y^2\mu_Y}\right]} \quad (3.5)$$

On substituting the value of k_{opt} in equation (3.4) we get the minimum value of MSE

$$MSE(\bar{y}_k)_{min} = \frac{\sigma_Y^2}{n}(1-\rho^2) + \frac{1}{n}\left[\sigma_u^2 + \rho 2\sigma Y 2\sigma v 2\sigma X 2 - \rho\mu 1200 CY\sigma Y + 2\sigma u 2 + \sigma Y 2 - \mu 0300\sigma YCY - 2\rho 2\sigma X\sigma Y 2n\mu Y 24CY 2\theta Y + AY - 4\mu 0300\sigma Y 2\mu Y\right] \quad (3.6)$$

4. Theoretical Efficiency Comparison

4.1 Proposed Estimator versus usual Linear Regression Estimator

From Maneesha and Singh (2002),MSE of usual linear regression estimator in presence of measurement errors is given as,

$$MSE(\bar{y}_{lr}) = \frac{\sigma_Y^2}{n}(1-\rho^2) + \frac{1}{n}\left[\sigma_u^2 + \rho^2\frac{\sigma_Y^2}{\sigma_X^2}\sigma_v^2\right]$$

MSE of proposed estimator is given as,

$$MSE(\bar{y}_k)_{min} = \frac{\sigma_Y^2}{n}(1-\rho^2) + \frac{1}{n}\left[\sigma_u^2 + \rho^2\sigma_Y^2\left(\frac{\sigma_v^2}{\sigma_X^2}\right)\right] - \frac{\left[\frac{\rho\mu_{1200}}{C_Y\sigma_Y} + 2(\sigma_u^2+\sigma_Y^2) - \frac{\mu_{0300}}{\sigma_Y C_Y} - 2\rho^2\sigma_X\sigma_Y\right]^2}{n\mu_Y^2\left[\frac{4C_Y^2}{\theta_Y} + A_Y - 4\frac{\mu_{0300}}{\sigma_Y^2\mu_Y}\right]}$$

Proposed estimator $\bar{y}_k$ will be more efficient than that of usual linear regression estimator $\bar{y}_{lr}$ if

$$MSE(\bar{y}_{lr}) - MSE(\bar{y}_k)_{min} > 0$$

$$\frac{\left[\frac{\rho\mu_{1200}}{C_Y\sigma_Y} + 2(\sigma_u^2+\sigma_Y^2) - \frac{\mu_{0300}}{\sigma_Y C_Y} - 2\rho^2\sigma_X\sigma_Y\right]^2}{n\mu_Y^2\left[\frac{4C_Y^2}{\theta_Y} + A_Y - 4\frac{\mu_{0300}}{\sigma_Y^2\mu_Y}\right]} > 0 \quad (4.1)$$

Thus the proposed estimator $\bar{y}_k$ will be more efficient than the usual linear regression estimator $\bar{y}_{lr}$ if the condition (4.1) is satisfied by the data set.

5. Simulation Study

We demonstrate the performance of all estimators by generating a sample from Normal distribution by using R software. The auxiliary information on variable X has been generated by N (5,10) population. This type of population is very relevant in most socio -economic situations with one interest and one auxiliary variable. 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, $\mu_X = 4.95$, $\mu_Y = 4.93$, $\sigma_X^2 = 99.38$, $\sigma_Y^2 = 100.12$, $\sigma_u^2 = 25.57$, $\sigma_v^2 = 24.28$, $\rho_{XY} = 0.99$ C_X = 2.012, C_Y = 2.029, λ = - 0.038 , A_X = 3.05, A_Y = 3.11 ,By using these values, the mean squared error (MSE) of the estimators of our interest are given in the following table

Table 1: MSEs of usual linear regression estimator and proposed estimator with and without measurement errors

Estimators	$\bar{y}_{lr}$	$\bar{y}_k$
MSE With measurement errors	3.98	2.912
MSE Without measurements	0.0001143	0.0000105

6. Conclusions

- When the observations are subject to measurement errors, condition (4.1) is satisfied by the data set. So our proposed estimator $\bar{y}_k$ is more efficient than that of usual linear regression estimator for this data set in the sense of having lesser MSE than the usual linear regression estimator.
- If observations are free from measurement errors then the proposed estimator is more efficient than the usual linear regression estimators.
- The percent relative efficiency (PRE) of the proposed estimator over the usual linear regression estimator in presence of measurement errors is 136%.
- The percent relative efficiency (PRE) of the proposed estimator over the usual linear regression estimator without measurement errors is 1088%.

References

[1] W.G. Cochran, Sampling Techniques, Second Edition, Wiley Eastern Private Limited, New Delhi, 1963.
[2] W.G. Cochran, "Errors of measurement in statistics" ,Technometrics, 10, 637-666, 1968.
[3] Judith. Lessler, T. and William, D. Kalsbeek, Non Sampling Error in Surveys, John Wiley and Sons. Inc, 1992.
[4] P.V. Sukhatme, B.V. Sukhatme, S. ,Sukhatme .and Ashok ,Sampling Theory of Surveys with Applications, Iowa State University Press, Ams(USA). and Indian Society of Agricultural Statistics, New Delhi (India), 1984.
[5] Hansen, Moris H. William N. H., Etes. Marks,And W. Parker M., "Response Errors in Surveys", JASA, 46, 147-190, 1951.
[6] Hansen, Morris H, William N. Hurwitz and Max A. Bershad "Measurement Errors in Survey", JASA, 46, 147-190, 1961.
[7] P.C.Mahalanobis,"Recent Experiments in Statistical Sampling in the Indian Statistical Institute", JRSS,109,327-328,1946.
[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, Vol.16, No.5, 821-840, 2009.
[9] S. Kumar, S. Bhougal, N.S Nataraja," Estimation of Population Mean in Presence of Non Response and Measurement Error Error", Revista Colombiana de Estadistica , **38**,145-161,2015.
[10] Maneesha and R. Karan Singh, "An Estimation of Population Mean in the Presence of Measurement Errors, Jour. Ind..Soc.Ag. Statistics",vol 54, 2001.
[11] Maneesha and Singh R.Karan "Role of Regression Estimator Involving Measurement Errors", Brazilian Journal of Probability and Statistics, Vol 16, pp 39-46, 2002.

[12] Salabh, "Ratio Method of Estimation in the Presence of Measurement Errors", Jour. Ind. Soc. Ag. Statistics, Vol .I, No-2, 150-155, 1997.
[13] Ph.D thesis, "On Some Classes of Estimators in Sampling Theory Using Auxiliary Information" by Rachana Maithani submitted and awarded to the Department of Statistics, University of Lucknow, 2013.
[14] P. Maiti, "Estimation of Non Sampling Variance Components under the Linear Model Approach", Statistics in Transition – new series, Vol. 10, no.2 , pp.193-222, 2009.
[15] J. Allen, H.P. Singh and F. Smarandache, "A family of estimators of population mean using multiauxiliary information in presence of measurement errors",International Journal of Social Economics, 30, 837-849, 2003.
[16] Shalabh, "Predictions of values of variables in linear measurement error model", Journal of Applied Statistics, 27, 475-482, 2000.
[17] H.P. Singh, and N. Karpe, "Ratio Product estimator for population mean in presence of measurement errors", Journal of Applied Statistical Science, 16, 49-64, 2008 a.
[18] H.P. Singh, and N. Karpe, "Estimation of population Variance using auxiliary information in the presence of Measurement errors", Statistics in Transition, 9, 443-470. 2008b.
[19] H.P. Singh, and N Karpe,"A class of estimators using auxiliary information for estimating finite population variance in nce of presence of measurement errors", Communication in Statistics-Theory and Methods, 38, 734-741, 2009.
[20] A.K. Srivastava, and Shalabh, "Asymptotic efficiency properties of least square in an ultrastructural model Test", 6, 419- 31, 1997.