Journal of Modern Applied Statistical Methods

Volume 12 | Issue 2 Article 13

11-1-2013

A Generalized Class of Estimators for Finite Population Variance in Presence of Measurement Errors

Prayas Sharma
Banaras Hindu University, Varanasi, India, prayassharma02@gmail.com

Rajesh Singh
Banaras Hindu University, Varanasi, India, rsinghstat@gmail.com

Journal of Modern Applied Statistical Methods
November 2013, Vol. 12, No. 2, 231-241.

ISSN 1538 − 9472

A Generalized Class of Estimators for Finite Population Variance in Presence of Measurement Errors

Prayas Sharma
Banaras Hindu University
Varanasi, India

Rajesh Singh
Banaras Hindu University
Varanasi, India

The problem of estimating the population variance is presented using auxiliary information in the presence of measurement errors. The estimators in this article use auxiliary information to improve efficiency and assume that measurement error is present both in study and auxiliary variable. A numerical study is carried out to compare the performance of the proposed estimator with other estimators and the variance per unit estimator in the presence of measurement errors.

Keywords: Population mean, study variate, auxiliary variates, mean squared error, measurement errors, efficiency.

Introduction

Over the past several decades, statisticians are paying their attention towards the problem of estimation of parameters in the presence of measurement errors. In survey sampling, the properties of estimators based on data usually presuppose that the observations are the correct measurements on characteristics being studied. However, this assumption is not satisfied in many applications and data is contaminated with measurement errors, such as non-response errors, reporting errors, and computing errors. These measurement errors make the result invalid, which are meant for no measurement error case. If measurement errors are very small and we can neglect it, then the statistical inferences based on observed data continue to remain valid. On the contrary, when they are not appreciably small and negligible, the inferences may not be simply invalid and inaccurate but may often lead to unexpected, undesirable and unfortunate consequences (see Srivastava and Shalabh, 2001). Some important sources of measurement errors in

Prayas Sharma is a Research Fellow in the Department of Statistics. Email him at prayassharma02@gmail.com. Rajesh Singh is Assistant professor in the Department of Statistics. Email at: rsinghstat@gmail.com.

survey data are discussed in Cochran (1968), Shalabh (1997), and Sud and Srivastva (2000). Singh and Karpe (2008, 2010), Kumar et al. (2011a, b) studied some estimators of population mean under measurement error.

Many authors, including Das and Tripathi (1978), Srivastava and Jhajj (1980), Singh and Karpe (2009) and Diana and Giordan (2012), studied the estimation of population Variance σ_y^2 of the study variable y using auxiliary information in the presence of measurement errors. The problem of estimating the population variance and its properties are studied here in the presence of measurement errors.

Consider a finite population $U = (U_1, U_2, \ldots\ldots U_N)$ of N units. Let Y and X be the study variate and auxiliary variate, respectively. Suppose a set of n paired observations are obtained through simple random sampling procedure on two characteristics X and Y. Further assume that x_i and y_i for the i^{th} sampling units are observed with measurement error as opposed to their true values (X_i, Y_i) For a simple random sampling scheme, let (x_i, y_i) be observed values instead of the true values (X_i, Y_i) for i^{th} (i=1.2….n) unit, as

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

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

where u_i and v_i are associated measurement errors which are stochastic in nature with mean zero and variances σ_u^2 and σ_v^2, respectively. Further, let the u_i's and v_i's are uncorrelated although X_i's and Y_i's are correlated .

Let the population means of X and Y characteristics be μ_x and μ_y , population variances of (x, y) be (σ_x^2, σ_y^2) and let ρ be the population correlation coefficient between x and y respectively (see Manisha and Singh (2002)).

Notations

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 estimator of population means $\bar{X}$ and $\bar{Y}$, respectively but $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ and $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$ are not unbiased estimator of (σ_x^2, σ_y^2), respectively. The expected values of s_x^2 and s_y^2 in the presence of measurement error are, given by,

$$E\left(s_x^2\right) = \sigma_x^2 + \sigma_v^2$$
$$E\left(s_y^2\right) = \sigma_y^2 + \sigma_u^2$$

When the error variance σ_v^2 is known, the unbiased estimator of σ_x^2, is $\hat{\sigma}_x^2 = s_x^2 - \sigma_v^2 > 0$, and when σ_u^2 is known, then the unbiased estimator of σ_y^2 is $\hat{\sigma}_y^2 = s_y^2 - \sigma_u^2 > 0$.

Define

$$\hat{\sigma}_y^2 = \sigma_y^2\left(1+e_0\right)$$
$$\bar{x} = \mu_x\left(1+e_1\right)$$

such that

$$E\left(e_0\right) = E\left(e_1\right) = 0,$$
$$E(e_1^2) = \frac{C_x^2}{n}\left(1+\frac{\sigma_v^2}{\sigma_x^2}\right) = \frac{C_x^2}{n\theta_x},$$

and to the first degree of approximation (when finite population correction factor is ignored)

$$E\left(e_0^2\right) = \frac{A_y}{n}, \; E\left(e_0 e_1\right) = \frac{\lambda C_x}{n}.$$

where,

$$A_y = \left\{\gamma_{2y} + \gamma_{2u}\frac{\sigma_u^4}{\sigma_y^4} + 2\left(1+\frac{\sigma_u^2}{\sigma_y^2}\right)^2\right\} \;, \quad \lambda = \frac{\mu_{12}(x,y)}{\sigma_x\sigma_y^2} \;, \quad C_x = \frac{\sigma_x}{\mu_x} \;, \quad \theta_x = \frac{\sigma_x^2}{\sigma_x^2+\sigma_v^2} \;,$$
$$\theta_y = \frac{\sigma_y^2}{\sigma_y^2+\sigma_u^2} \;, \; \gamma_{2y} = \beta_2(y)-3 \;, \; \gamma_{2u} = \beta_2(u)-3 \,, \; \beta_2(u) = \frac{\mu_4(u)}{\mu_2^2(u)} \,, \; \beta_2(y) = \frac{\mu_4(y)}{\mu_2^2(y)} \;,$$
$$\mu_4(y) = E(Y_i-\mu_y)^4 \;, \; \mu_4(u) = E(u_i^4) \,.$$

θ_x and θ_y are the reliability ratios of X and Y, respectively, lying between 0 and 1.

Estimator of population variance under measurement error

According to Koyuncu and Kadilar (2010), a regression type estimator t_1 is defined as

$$t_1 = w_1\hat{\sigma}_y^2 + w_2(\mu_x - \overline{x}) \tag{3}$$

where w_1 and w_2 are constants that have no restriction .

Expression (3) can be written as

$$t_1 - \sigma_y^2 = (w_1 - 1)\sigma_y^2 + w_1\sigma_y^2 e_0 - w_2\mu_x e_1 \tag{4}$$

Taking expectation both sides of (4), results in

$$Bias(t_1) = \sigma_y^2(w_1 - 1) \tag{5}$$

Squaring both sides of (4)

$$\left(t_1 - \sigma_y^2\right)^2 = \left[\left(w_1 - 1\right)\sigma_y^2 + w_1\sigma_y^2 e_0 - w_2\mu_x e_1\right] \tag{6}$$

or

$$\begin{aligned}\left(t_1 - \sigma_y^2\right)^2 = \Big[&\sigma_y^4(w_1-1)^2 + w_1^2\sigma_y^4 e_0^2 + w_2^2\mu_x^2 e_1^2 + 2(w_1-1)w_1\sigma_y^4 e_0 \\ &-2(w_1-1)w_2\sigma_y^2\mu_x e_1 - 2w_1 w_2\sigma_y^2\mu_x e_0 e_1)\Big]\end{aligned} \tag{7}$$

Simplifying equation (7), taking expectations and using notations, results in the mean square error of t_1 up to first order of approximation, as

$$MSE(t_1) = \left[\sigma_y^4 w_1^2(\frac{A_y}{n} + 1) + (1 - 2w_1)\sigma_y^4 + w_2^2\mu_x^2\frac{C_x^2}{n\theta_x} - \frac{2w_1w_2\mu_x\sigma_y^2\lambda C_x}{n}\right] \tag{8}$$

In the case, when the measurement error is zero, *MSE* of t_1 without measurement error is given by,

$$MSE^*(t_1) = \frac{\sigma_y^4}{n}\left\{\gamma_{2y} + 2 + n\right\} + (1-2w_1)\sigma_y^4 + w_2^2\mu_x^2\frac{C_x^2}{n} - 2w_1w_2\mu_x\sigma_y^2\lambda\frac{C_x}{n}\Bigg] \tag{9}$$

and

$$M_{t_1} = \frac{\sigma_y^2}{n}\left[\frac{\sigma_u^4}{\sigma_y^4}\gamma_{2u} + 2\left(\frac{\sigma_u^4}{\sigma_y^4}\right)^2 + 4\frac{\sigma_u^4}{\sigma_y^4}\right] + w_2^2\mu_x^2\frac{C_x^2}{n}\frac{\sigma_v^2}{\sigma_x^2} \tag{10}$$

is the contribution of measurement errors in the *MSE* of estimator t_1.

Differentiating (8) with respect to w_1 and w_2 partially, equating them to zero and after simplification, results in the optimum values of w_1 and w_2, respectively as

$$w_1^* = \frac{-\sigma_y^4 B}{C^2 - AB}, \; w_2^* = \frac{-\sigma_y^4 C}{C^2 - AB} \tag{11}$$

where, $A = (\frac{A_y}{n} + 1)\sigma_y^4$, $B = \frac{\mu_x^2 C_x^2}{n\theta_x}$ and $C = \frac{\sigma_y^2 \mu_x C_x \lambda}{n}$.

Using the values of ω_1^* and ω_2^* from equation (11) into equation (8), gives the minimum *MSE* of the estimator t_2 in terms of *A*, *B* and *C* as

$$MSE(t_1)_{\min} = \left(\frac{\sigma_y^4}{(C^2 - AB)}\right)^2\left[\frac{\left(C^2 - AB\right)^2}{\sigma_y^4} + 3BC^2 - AB^2 - 2BC\right] \tag{12}$$

Another estimator under measurement error

Based on Solanki and Singh (2012), an estimator t_3 is defined as

$$t_2 = \hat{\sigma}_y^2 \left\{ 2\text{-}\left(\frac{\bar{x}}{\mu_x}\right)^{\alpha} \exp\left[\frac{\beta\left(\bar{x}-\mu_x\right)}{\left(\bar{x}+\mu_x\right)}\right] \right\} \tag{13}$$

where α and β are suitably chosen constants.
Expressing the estimator t_2 , in terms of e's is

$$t_2 = \hat{\sigma}_y^2 \left[2-\left(1+e_1\right)^{\alpha} \exp\left\{ (\frac{\beta e_1}{2})\left(1+\frac{e_1}{2}\right)^{-1} \right\} \right] \tag{14}$$

Expanding equation (14) and simplifying results in

$$\left(t_2-\sigma_y^2\right) = \sigma_y^2 \left[e_0 - \frac{k}{2}(e_1+e_0e_1) - \frac{e_1^2}{8}\left(k^2-2k\right) \right] \tag{15}$$

where $k = \left(\beta + 2\alpha\right)$.

On taking expectations of both sides of (15), the bias of the estimator t_3 up to the first order of approximation is obtained as

$$Bias\left(t_2\right) = \sigma_y^2 \left[\text{-}\frac{\text{k}}{2}\frac{\lambda C_x}{n} - \left(\frac{k^2-2k}{8}\right)\frac{C_x^2}{n\theta_x} \right] \tag{16}$$

Squaring both sides of (15) and after simplification,

$$\left(t_2-\sigma_y^2\right)^2 = \sigma_y^4 \left[e_0 + \frac{k^2}{4}e_1^2 - ke_0e_1 \right] \tag{17}$$

Taking expectations of (17) and using notations, the *MSE* of estimator t_2 is calculated as

$$MSE(t_2) = \frac{\sigma_y^4}{n\theta_x}\left[A_y\theta_x + \frac{k^2}{4}C_x^2 - k\lambda C_x\theta_x \right] \tag{18}$$

Differentiating equation (18) with respect to k and equating to zero and after simplification the optimum value of k is

$$k^* = 2\frac{\lambda\theta_x}{C_x} \tag{19}$$

Putting the optimum value of k from (19) to (18), results in the minimum *MSE* of estimator t_2 as

$$MSE(t_2)_{\min} = \frac{\sigma_y^4}{n}\left[A_y - \lambda^2\theta_x\right] \tag{20}$$

Remark:

Singh and Karpe (2009) defined a class of estimator for σ_y^2 as

$$t_d = \hat{\sigma}_y^2 d(b) \tag{21}$$

where, $d(b)$ is a function of b such that $d(1)= 1$, and certain other conditions, similar to those given in Srivastava (1971). The minimum *MSE* of t_d is given by,

$$MSE(t_d)_{\min} = \frac{\sigma_y^4}{n}\left[A_y - \lambda^2\theta_x\right] \tag{22}$$

which is the same as the minimum *MSE* of estimator t_2, given in equation (20).

A General Class of Estimators

A general class of estimator t_3 is proposed as

$$t_3 = \left[m_1\hat{\sigma}_y^2 + m_2\left(\mu_x - \bar{x}\right)\right]\left\{2\text{-}\left(\frac{\bar{x}}{\mu_x}\right)^{\alpha}\exp\left[\frac{\beta\left(\bar{x}-\mu_x\right)}{\left(\bar{x}+\mu_x\right)}\right]\right\} \tag{23}$$

Where m_1 and m_2 are constants chosen so as to minimize the mean squared error of the estimator t_3.

Equation (23) can be expressed in terms of e's as

$$t_3 = \left[m_1\sigma_y^2 + m_1\sigma_y^2 e_0 - m_2\mu_x e_1\right]\left[1 - \frac{k}{2}e_1 - \frac{\left(k^2-2k\right)}{8}e_1^2\right] \tag{24}$$

Expanding equation (24) and subtracting σ_y^2 from both sides, results in

$$\left(t_3-\sigma_y^2\right) = \Big[\left(m_1-1\right)\sigma_y^2 - \frac{k}{2}m_1\sigma_y^2 e_1 + m_1\sigma_y^2 e_0 - m_2\mu_x e_1 - \frac{e_1^2}{8}\sigma_y^2 m_1\left(k^2-2k\right) - \frac{\sigma_y^2 m_1 k}{2}e_0 e_1 + \frac{k}{2}m_2\mu_x e_1^2\Big] \tag{25}$$

On taking expectations of both sides of (25) the bias of the estimator t_3 up to the first order approximation is obtained as

$$Bias\left(t_3\right) = \left(m_1-1\right)\sigma_y^2 - \frac{1}{8}\sigma_y^2 m_1\left(k^2-2k\right)\frac{C_x^2}{n\theta_x} - \frac{\sigma_y^2 m_1 k}{2}\frac{\lambda C_x}{n} + \frac{k}{2}m_2\mu_x\frac{C_x^2}{n\theta_x} \tag{26}$$

Squaring both sides of (25), results in

$$\left(t_3-\sigma_y^2\right)^2 = \left[\left(m_1-1\right)\sigma_y^2 - \frac{k}{2}m_1\sigma_y^2 e_1 + m_1\sigma_y^2 e_0 - m_2\mu_x e_1\right]^2 \tag{27}$$

Simplifying equation (27) and taking expectations both sides the *MSE* of estimator t_3 up to the first order of approximation is obtained as

$$MSE(t_3) = \left[\left(1-2m_1\right)\sigma_y^4 + m_1^2 P + m_2^2 Q - m_1 m_2 R\right] \tag{28}$$

where $P = \left(1 + \frac{A_y}{n} + \frac{k^2C_x^2}{4n\theta_x} - \frac{k}{n}\lambda C_x\right)\sigma_y^4$, $Q = \frac{\mu_x^2 C_x^2}{n\theta_x}$ and $R = \sigma_y^2\left(k\frac{C_x^2}{\theta_x} + 2\lambda C_x\right)\frac{\mu_x}{n}$.

Minimizing *MSE* t_3 with respect to m_1 and m_2 the optimum values of m_1 and m_2 is

$$\mathrm{m}_1^* = \frac{-4Q\sigma_y^4}{R^2 - 4PQ} \quad \text{and} \quad m_2^* = \frac{-2R\sigma_y^4}{R^2 - 4PQ}$$

Putting the optimum values of m_1 and m_2 in equation (28) results in the minimum *MSE* of estimator t_3 as

$$MSE\left(t_3\right) = \sigma_y^4\left[1 = \frac{4\sigma_y^4 Q}{\left(4PQ - R^2\right)}\right] \tag{29}$$

Empirical Study

Data Statistics:

The data used for empirical study was taken from Gujrati and Sangeetha (2007) -pg, 539., where,

Y_i = True consumption expenditure,
X_i = True income,
y_i = Measured consumption expenditure,
x_i = Measured income.

From the data given we get the following parameter values:

Table 1. Parameter values from empirical data

N	μ_y	μ_x	σ_y^2	σ_x^2	ρ	σ_u^2	σ_v^2
10	127	170	1278	3300	0.964	36.0	36.0

Table 2. Showing the *MSE* of the estimators with and without measurement errors

Estimators	***MSE* without meas. Error**	**Contribution of meas. Errors in *MSE***	***MSE* with meas. Errors**
$\hat{\sigma}_y^2$	245670	35458	281128
t_1	229734	30354	260088

Table 2 continued.

Estimators	***MSE* without meas. Error**	**Contribution of meas. Errors in *MSE***	***MSE* with meas. Errors**
$t_{2\min}$	245411	35461	280872
$t_{3\min}(\alpha=1,\beta=0)$	247440	30442	277862
$(\alpha=0,\beta=1)$	234402	30555	267957
$(\alpha=1,\beta=1)$	268144	30219	298363
$(\alpha=1,\beta=-1)$	234402	33555	267957
$(\alpha=0,\beta=-1)$	231969	30600	262569
$(\alpha=-0.9,\beta=2)$	**229145**	**30365**	**259510**

Conclusion

Table 2 shows that the *MSE* of proposed estimator t_3 (for $\alpha=-0.9,\beta=2$) is minimum among all other estimators considered. It is also observed that the effect due to measurement error on the estimator t_1 and usual estimators is less than the effect on the estimator t_2 under measurement error for this given data set.

References

Allen, J., Singh, H. P., & Smarandache, F. (2003). A family of estimators Of population mean using multi auxiliary information in presence of measurement errors. *International Journal of Social Economics 30*(7), 837–849.

Cochran, W. G. (1968). Errors of Measurement in statistics. *Technometrics 10,* 637-666

Das, A. K., & Tripathi, T. P. (1978). Use of auxiliary information in estimating the Finite population variance. *Sankhya C 4*, 139 - 148

Diana, G., & Giordan, M. (2012). Finite Population Variance Estimation in Presence of Measurement Errors. *Communication in Statistics Theory and Methods, 41*, 4302-4314.

Gujarati, D. N., & Sangeetha (2007). *Basic econometrics*. McGraw – Hill.

Koyuncu, N., & Kadilar, C. (2010). On the family of estimators of Population mean in stratified sampling.Pakistan *Journal of Statistics, 26,* 427-443.

Kumar, M., Singh, R., Singh, A. K., & Smarandache, F. (2011a). Some ratio Type estimators under measurement errors. *World Applied Sciences Journal, 14*(2), 272 - 276.

Kumar, M., Singh, R., Sawan, N., & Chauhan, P. (2011b). Exponential ratio method Of estimators in the presence of measurement errors. *International Journal of Agricultural and Statistical Sciences 7*(2), 457-461.

Manisha, M., & Singh, R. K. (2002). Role of regression estimator involving Measurement errors. Brazilian *Journal of Probability and Statistics 16*, 39- 46.

Shalabh. (1997). Ratio method of estimation in the presence of measurement errors. *Journal of Indian Society of Agricultural Statistics 50*(2), 150– 155.

Singh, H. P. & Karpe, N. (2008). Ratio product estimator for population mean in presence of measurement errors. *Journal of Applied Statistical Sciences, 16*(4), 49-64.

Singh, H. P. & Karpe, N. (2009). Class of estimators using auxiliary Information for estimating finite population variance in presence of measurement errors. *Communication in Statistics Theory and Methods, 38*, 734-741.

Singh, H. P. & Karpe, N. (2010). Effect of measurement errors on the Separate And combined ratio and product estimators in Stratified random sampling. *Journal of Modern Applied Statistical Methods, 9*(2), 338-402.

Solanki R., Singh H.P., & Rathour A. (2012). An alternative estimator for estimating the finite population mean using auxiliary information in sample surveys. *ISRN Probability and Statistics*, doi:10.5402/2012/657682.

Srivastava, M. S. (1971). On Fixed-Width Confidence Bounds for Regression Parameters, *Annals of Mathematical Statistics, 42*, 1403-1411.

Srivastava, A., K., & Shalabh. (2001). Effect of Measurement Errors On the Regression Method of Estimation in Survey Sampling. *Journal of Statistical Research*, *35*(2), 35-44.

Srivastava, S. K., & Jhajj, H.S. (1980) A class of estimators using auxiliary information for estimating finite population variance. *Sankhya Ser. C 42*, 87-96.

Sud, U. C., & Srivastava, S. K. (2000). Estimation of population mean in repeat surveys in the presence of measurement errors. *Journal of the Indian Society of Agricultural Statistics, 53*(2), 125-133.