SHORT COMMUNICATIONS

Arab J. Math. Sc.
Volume 5, Number 1, June 1999
pp. 51-55

A NOTE ON THE ACCURACY OF SOME ESTIMATION TECHNIQUES IN THE PRESENCE OF MEASUREMENT ERRORS

L. N. SAHOO AND R.K. SAHOO

ABSTRACT. When the observed values for the sample elements are equipped with measurement errors, the precision of the survey estimate is adversely affected. Recently, Shalabh (1997) examined various properties of ratio estimate in the presence of measurement errors. In this note, we study the effect of measurement errors on the accuracy of some standard survey estimates designed to make efficient use of auxiliary information about the population.

1. INTRODUCTION

Consider a bivariate population U in $\begin{pmatrix} Y \\ X \end{pmatrix}$ with mean vector $\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}$ and dispersion matrix $\begin{pmatrix} \sigma_y^2 & \rho\sigma_y\sigma_x \\ \rho\sigma_y\sigma_x & \sigma_x^2 \end{pmatrix}$, where Y is the survey variable, X is an auxiliary variable supposed to be highly correlated with Y, and ρ is the correlation coefficient between them. It is assumed that for each element $k \in U$, there exist true values Y_k and X_k and that the objective is to estimate μ_y . Suppose that a sample s of size n is selected from U by simple random sampling with replacement. Our idea is to obtain true values $\begin{pmatrix} Y_k \\ X_k \end{pmatrix}$, $k \in s$, but what we actually obtain through the measurement procedure the observed $\begin{pmatrix} y_k \\ x_k \end{pmatrix}$, $k \in s$, are composed of

1991 Mathematics Subject Classification: 62D05.

the true values and the random measurement errors providing a sample mean vector denoted by $\begin{pmatrix} \overline{y} \\ \overline{x} \end{pmatrix}$. For lack of better values, the y_k and x_k must nevertheless be used in the computation of estimates. The vector of measurement errors associated with element $k \in s$ is now defined as

$$\begin{pmatrix} u_k \\ v_k \end{pmatrix} = \begin{pmatrix} y_k - Y_k \\ x_k - X_k \end{pmatrix},$$

which is assumed to be stochastic with

$$E\begin{pmatrix} u_k \\ v_k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} and V\begin{pmatrix} u_k \\ v_k \end{pmatrix} = \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}.$$

In this work we shall examine the consequence of the u_k's and v_k's on the precision of the following estimates:

Mean per unit estimate	:	$\overline{y}$
Ratio estimate	:	$\overline{y}_R = \overline{y}\mu_x/\overline{x}$
Product estimate	:	$\overline{y}_P = \overline{yx}/\mu_x$
Regression estimate	:	$\overline{y}_{RG} = \overline{y} - b(\overline{x} - \mu_x)$, b being the sample
	:	regression coefficient of Y on X.

2. MAIN RESULTS

The survey can be viewed as a two-stage process, with each stage contributing randomness. The first one being the sample selection which results in selecting the sample s. The second one being the measuring process that generates an observed value $y_k, k \in s$. Hence, we have,

$$\begin{aligned} E(.) &= E[E(./s)], \\ V(.) &= V[E(./s)] + E[V(./s)], \\ Cov(.,.) &= Cov[E(./s), E(./s)] + E[Cov(.,./s)], \end{aligned}$$

so that, omitting details of calculations we obtain,

$$E\begin{pmatrix} \overline{y} \\ \overline{x} \end{pmatrix} = \begin{pmatrix} \mu y \\ \mu x \end{pmatrix} \text{ and } V\begin{pmatrix} \overline{y} \\ \overline{x} \end{pmatrix} = \frac{1}{n}\begin{pmatrix} \sigma_y^2\sigma_u^2 & \rho\sigma_y\sigma_x \\ \rho\sigma_y\sigma_x & \sigma_x^2 + \sigma_x^2 \end{pmatrix}.$$

Thus, we now have

$$V(\bar{y}) = \frac{1}{n}\left(\sigma_y^2 + \sigma_u^2\right),$$

and adopting the usual procedure as described in many standard text books [cf. Cochran (1977)], we get the mean square errors of $\overline{y}_R, \overline{y}_P$ and $\overline{y}_{RG}$ up to order n^{-1} as

$$M(\overline{y}_R) = \frac{1}{n}\left[\left(\sigma_y^2 - 2R\rho\sigma_y\sigma_x + R^2\sigma_x^2\right) + \left(\sigma_u^2 + R^2\sigma_v^2\right)\right] \tag{2.1}$$

$$M(\overline{y}_P) = \frac{1}{n}\left[\left(\sigma_y^2 + 2R\rho\sigma_y\sigma_x + R^2\sigma_x^2\right) + \left(\sigma_u^2 + R^2\sigma_v^2\right)\right] \tag{2.2}$$

$$M(\overline{y}_{RG}) = \frac{1}{n}\left[\sigma_y^2(1-\rho^2) + \left(\sigma_u^2 + \beta^2\sigma_v^2\right)\right] \tag{2.3}$$

where $R = \mu_y/\mu_x$ and $\beta = \rho\sigma_y/\sigma_x$.

Examining the above expressions we observe that the precision in each case decreases when the measurement errors are present. These expressions now lead to the following easily verified results:

(i) $V(\overline{y}) \gtrless M(\overline{y}_R)$ if

$$\lambda \gtrless \frac{1}{2}\left(1 + \frac{\sigma_v^2}{\sigma_x^2}\right) \tag{2.4}$$

and $V(\overline{y}) \gtrless M(\overline{y}_P)$ if

$$\lambda \lessgtr -\frac{1}{2}\left(1 + \frac{\sigma_v^2}{\sigma_x^2}\right) \tag{2.5}$$

where $\lambda = \beta/R$. So, we find that $\overline{y}$ is superior to both $\overline{y}_R$ and $\overline{y}_P$ when

$$-\frac{1}{2}\left(1 + \frac{\sigma_v^2}{\sigma_x^2}\right) \leq \lambda \leq \frac{1}{2}\left(1 + \frac{\sigma_v^2}{\sigma_x^2}\right). \tag{2.6}$$

Thus, even in those situations where the ratio or product estimator is known to have better performance than mean per unit estimator in the absence of measurement errors in X-variable, cases may arise in which the former ones turn out to be less efficient than the later one in the presence of measurement errors. Since $\overline{y}$ is better than $\overline{y}_R$ and $\overline{y}_P$ for $-\frac{1}{2} \leq \lambda \leq \frac{1}{2}$, the inequalities in (2.7) clearly indicate that the presence of measurement errors in X-variable definitely widens the region of preference for $\overline{y}$ in terms of λ.

(ii) $V(\overline{y}) \gtrless M(\overline{y}_{RG})$

(2.7) $$\sigma_x^2 \gtrless \sigma_v^2.$$

Thus, if the auxiliary variable is so poorly measured that the error variance σ_v^2 is larger than σ_x^2, then $\overline{y}_{RG}$ is inferior to $\overline{y}$ although $\overline{y}$ is unconditionally inferior to $\bar{y}_{RG}$ when X-values are free from measurement errors.

(iii) $M(\bar{y}_R) \gtrless M(\bar{y}_{RG})$ if

(2.8) $$(1-\lambda)^2 + (1-\lambda^2)\frac{\sigma_v^2}{\sigma_x^2} \gtrless 0$$

and $M(\bar{y}_P) \gtrless M(\bar{y}_{RG})$ if

(2.9) $$(1+\lambda)^2 + (1-\lambda^2)\frac{\sigma_v^2}{\sigma_x^2} \gtrless 0.$$

Thus, even if $\bar{y}_{RG}$ is more precise than both $\bar{y}_R$ and $\bar{y}_P$, the presence of measurement errors in X-values limits its scope of use compared to $\overline{y}_R$ or $\bar{y}_P$.

3. CONCLUSION

The findings in the preceding section lead to a conclusion that the measurement errors in X-variable may alter the preference ordering of the estimators derived under the assumption of absence of measurement errors. However, the measurement errors in Y-variable have no role to play in this kind of preference ordering.

REFERENCES

1. W.G. Cochran, *Sampling Techniques* (Third Edition), John Wiley & Sons, 1977.

2. Shalabh, *Ratio Method of Estimation in the Presence of Measurement Errors,* , Journal of the Indian Society of Agricultural Statistics, **50**(1997), 150-155.

DEPARTMENT OF STATISTICS, UTKAL UNIVERSITY, BHUBANESWAR 751004, INDIA.

Date received December 15, 1998.