Jour. Ind. Soc. Ag. Statistics
53(2), 2000 : 125-133

Estimation of Population Mean in Repeat Surveys in the Presence of Measurement Errors

U.C. Sud and A.K. Srivastava
Indian Agricultural Statistics Research Institute,
New Delhi-110012
(Received : January, 1998)

SUMMARY

Estimators of current occasion, change and sum over two occasions have been developed when the variables are subject to measurement errors. Expressions for gain in the precision for these estimators have been obtained. As in the absence of measurement errors, the theory of repeat surveys provide gains when the variables are in error.

Key words : Repeat surveys, Measurement errors, Minimum variance unbiased estimator.

1. Introduction

Data collected through surveys are often assumed to be free of measurement or response errors. In reality the data may be contaminated with measurement errors. Such errors can distort the data in several ways. For instance, the very method of data collection can result in errors. Measurement errors may also be attributable to the respondents, interviewer or the way in which the questionnaire is framed. Sometimes the nature of the variable is such that it may cause errors of measurement. This may happen in case of qualitative variables. Examples are variables which pertain to intelligence, preference, tastes, etc. Another source of measurement error is when the variable is conceptually well defined but observations can be obtained on some closely related substitutes. Such a situation is encountered when one is collecting data to measure the economic status or the level of education of individuals. Measurement errors can result in serious misleading inferences (Biemer *et al.* [1]). In this article relevant theory has been developed for repeat surveys on two occasions when the variables are subject to measurement errors.

2. The Measurement Error Model and Estimation for Current Occasion

Let (x_i, y_i ; $i = 1, \dots N$) be the observed values of the characteristics on first and second occasion respectively, the corresponding true values being

(η_i, ω_i ; i = 1, ... N). A simple random sample of 'n' units is obtained on the first occasion. A random sub-sample of $m = n\lambda$ units is retained for use on the second occasion. An independent sample of $u = n - m = n\mu$ units is selected (unmatched with the first occasion). The measurement errors for the two occasions are defined as

$$\varepsilon_i = x_i - \eta_i$$

$$e_i = y_i - \omega_i \tag{2.1}$$

We assume that

$$E_\xi(\varepsilon_i/i) = 0 \; ; \; E_\xi(e_i/i) = 0$$

$$E_\xi(\varepsilon_i^2/i) = \delta_{i1}^2 \; ; \; E_\xi(e_i^2/i) = \delta_{i2}^2 \tag{2.2}$$

where

E_ξ stands for expectation under the measurement error model.

Further, we assume that the measurement errors pertaining to the matched portion of the sample on two occasions are correlated.

Let $\bar{x}$ and $\bar{y}$ respectively denote the population mean based on observed values on the two occasions, the corresponding true values being $\bar{\eta}$ and $\bar{\omega}$ respectively.

where

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \; ; \; \bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i \; ; \; \bar{\eta} = \frac{1}{N} \sum_{i=1}^{N} \eta_i \text{ and } \bar{\omega} = \frac{1}{N} \sum_{i=1}^{N} \omega_i$$

The parameter of interest is $\bar{\omega}$.

Let the population variances for the observed values, true values and measurement errors on the two occasions be denoted by $\sigma_{0i}^2, \sigma_{ti}^2$ and $\sigma_{\delta i}^2$ (i = 1, 2) respectively. Further, ρ and ρ_e respectively denote correlation coefficients between true values and measurement errors on the two occasions.

It may be seen that $\sigma_{0i}^2 = \sigma_{ti}^2 + \sigma_{\delta i}^2 \;\; \forall \; i = 1, 2$

We follow the approach of Yates [4] and Patterson [2] in developing minimum variance unbiased linear estimator (MVULE) of $\bar{\omega}$ as under :

Let the single prime indicates units common to two occasions and a double prime indicates the units selected independently.

2^{nd} occasion	$\bar{y}'$	$\bar{y}''$
$\bar{x}''$	$\bar{x}'$	1^{st} occasion

Consider the following linear estimator of $\bar{\omega}$

$$\hat{M}_2 = a\hat{\bar{y}}'' + b\hat{\bar{y}}' + c\hat{\bar{x}}' + d\hat{\bar{x}}''$$

where

$$\hat{\bar{y}}'' = \frac{1}{u}\sum_{i=1}^{u} y_i \,;\ \hat{\bar{y}}' = \frac{1}{m}\sum_{i=1}^{m} y_i \,;\ \hat{\bar{x}}'' = \frac{1}{u}\sum_{i=1}^{u} x_i \text{ and } \hat{\bar{x}}' = \frac{1}{m}\sum_{i=1}^{m} x_i$$

This will be unbiased if

$$EE_\xi(\hat{M}_2) = \bar{\omega}$$

This condition means $\hat{M}_2$ will have the form

$$\hat{M}_2 = d(\hat{\bar{x}}'' - \hat{\bar{x}}') + (1-b)\hat{\bar{y}}'' + b\hat{\bar{y}}'$$

If it is MVLUE, it must be uncorrelated with every zero function (Rao, [3]). Hence it is uncorrelated with $\hat{\bar{x}}' - \hat{\bar{x}}''$ and $\hat{\bar{y}}' - \hat{\bar{y}}''$. Thus

$$\text{Cov}(\hat{\bar{y}}', \hat{M}_2) = \text{Cov}(\hat{\bar{y}}'', \hat{M}_2) \tag{2.3}$$

and

$$\text{Cov}(\hat{\bar{x}}', \hat{M}_2) = \text{Cov}(\hat{\bar{x}}'', \hat{M}_2) \tag{2.4}$$

where

$$\text{Cov}(\hat{\bar{y}}', \hat{M}_2) = E\,\text{Cov}_\xi(\hat{\bar{y}}', \hat{M}_2) + \text{Cov}\,E_\xi(\hat{\bar{y}}', \hat{M}_2)$$

other covariance terms in (2.3) and (2.4) can be similarly defined.

It can be shown that

$$\text{Cov}(\hat{\bar{y}}', \hat{M}_2) = \frac{b}{n\lambda}(\sigma_{t2}^2 + \sigma_{\delta 2}^2) - \frac{d}{n\lambda}(\rho\sigma_{t1}\sigma_{t2} + \rho_e\sigma_{\delta 1}\sigma_{\delta 2})$$

$$\text{Cov}(\hat{\bar{y}}'', \hat{M}_2) = \frac{(1-b)}{n\mu}(\sigma_{t2}^2 + \sigma_{\delta 2}^2)\,;\ \text{Cov}(\hat{\bar{x}}'', \hat{M}_2) = \frac{d}{n\mu}(\sigma_{t1}^2 + \sigma_{\delta 1}^2)$$

and

$$\text{Cov}(\hat{\bar{x}}', \hat{M}_2) = \frac{d}{n\lambda}(\sigma_{t1}^2 + \sigma_{\delta 1}^2) + \frac{b}{n\lambda}(\rho\sigma_{t1}\sigma_{t2} + \rho_e\sigma_{\delta 1}\sigma_{\delta 2})$$

Substituting these expressions in (2.3) and (2.4) we obtain

$$\frac{b}{n\lambda}(\sigma_{t2}^2+\sigma_{\delta 2}^2)-\frac{d}{n\lambda}(\rho\sigma_{t1}\sigma_{t2}+\rho_e\sigma_{\delta 1}\sigma_{\delta 2}) = \frac{(1-b)}{n\mu}(\sigma_{t2}^2+\sigma_{\delta 2}^2)$$

$$\frac{d}{n\mu}(\sigma_{t1}^2+\sigma_{\delta 1}^2) = -\frac{d}{n\lambda}(\sigma_{t1}^2+\sigma_{\delta 1}^2)+\frac{b}{n\lambda}(\rho\sigma_{t1}\sigma_{t2}+\rho_e\sigma_{\delta 1}\sigma_{\delta 2})$$

Solving the above equations for b and d, we obtain

$$b = \frac{\lambda}{1-\mu^2\rho_0^2} \text{ and } d = \frac{\lambda\mu\rho_o}{(1-\mu^2\rho_0^2)}\frac{\sigma_{02}}{\sigma_{01}}$$

where

$$\rho_0 = \frac{\rho\sigma_{t1}\sigma_{t2}+\rho_e\sigma_{\delta 1}\sigma_{\delta 2}}{\sigma_{01}\sigma_{02}}$$

Substituting the values of b and d in (2.2), the best linear estimator of $\bar{\omega}$ is given by

$$\hat{M}_2 = \frac{1}{1-\mu^2\rho_0^2}\left[\lambda\mu\rho\frac{\sigma_{02}}{\sigma_{01}}(\hat{\bar{x}}''-\hat{\bar{x}}')+\mu(1-\mu\rho_o^2)\hat{\bar{y}}''\right] \tag{2.5}$$

Now the variance of $\hat{M}_2$ equals the covariance of $\hat{M}_2$ and any unbiased estimator of $\bar{\omega}$.

Thus,

$$V(\hat{M}_2) = \text{Cov}(\hat{\bar{y}}'', \hat{M}_2)$$

$$= \frac{\sigma_{02}^2}{n}\frac{1-\mu\rho_o^2}{(1-\mu^2\rho_0^2)} \tag{2.6}$$

The optimum values of μ and λ in the sense of minimum variance of $\hat{M}_2$ equals

$$\mu = [1+(1-\rho_o^2)^{1/2}]^{-1} \text{ and } \lambda = (1-\rho_0^2)^{1/2}[1+(1-\rho_0^2)^{1/2}]^{-1} \tag{2.7}$$

Substituting the values of μ and λ in (2.6) we get the minimum variance of $\hat{M}_2$ as

$$V_{min}(\hat{M}_2) = \frac{\sigma_{02}^2}{2n\mu} \tag{2.8}$$

If a completely independent sample is taken on the second occasion the resulting estimator is given by

$$\hat{M}'_2 = \frac{1}{n} \sum_{i=1}^{n} y_i$$

The variance of $\hat{M}'_2$ can be seen to be equal to

$$V(\hat{M}'_2) = \frac{\sigma_{02}^2}{n} \tag{2.9}$$

Therefore, the percent gain in precision of $\hat{M}_2$ over $\hat{M}'_2$ is given by

$$\frac{1 - \sqrt{1 - \rho_0^2}}{1 + \sqrt{1 - \rho_0^2}} \times 100 \tag{2.10}$$

It is clear from expression (2.10) that the gain in precision is 0 for $\rho_0 = 0$ and 100% for $\rho_0 = 1$.

3. *Estimation of Change*

As in the case of $\hat{M}_2$, the best linear unbiased estimator of $\overline{\eta}$ will be of the form

$$\hat{M}_1 = \frac{1}{1 - \mu^2\rho_0^2}\left[\lambda\mu\rho_0 \frac{\sigma_{01}}{\sigma_{02}} (\hat{\bar{y}}'' - \hat{\bar{y}}') + \lambda \hat{\bar{x}}' + \mu (1 - \mu\rho_0^2) \hat{\bar{x}}''\right]$$

Thus, the best estimator of change $\Omega = \overline{\omega} - \overline{\eta}$ is given by

$$\hat{\Omega} = \hat{M}_2 - \hat{M}_1$$

$$= \frac{1}{1 - \mu^2\rho_0^2}\left[\left\{\lambda\mu\rho_0 \frac{\sigma_{02}}{\sigma_{01}} (\hat{\bar{x}}'' - \hat{\bar{x}}') + \mu (1 - \mu\rho_o^2) \hat{\bar{y}}'' + \lambda\hat{\bar{y}}'\right\} - \left\{\lambda\mu\rho_0 \frac{\sigma_{01}}{\sigma_{02}} (\hat{\bar{y}}'' - \hat{\bar{y}}') + \mu (1 - \mu\rho_o^2) \hat{\bar{x}}'' + \lambda\hat{\bar{x}}'\right\}\right] \tag{3.1}$$

Since $\hat{\bar{y}}'' - \hat{\bar{x}}''$ is an unbiased estimator of Ω, we have

$$V(\hat{\Omega}) = \text{Cov}(\hat{\bar{y}}'' - \hat{\bar{x}}'', \hat{\Omega})$$

$$= \frac{1}{n(1 - M^2\rho_0^2)} [(1 - \mu\rho_o^2)(\sigma_{02}^2 + \sigma_{01}^2) - 2\lambda\rho_0\sigma_{02}\sigma_{01}] \tag{3.2}$$

A simplified expression for the variance of change can be obtained if one assumes that $\sigma_{01}^2 = \sigma_{02}^2 = \sigma_0^2$. Under this assumption the expression for the variance of change is given by

$$V(\hat{\Omega}) = \frac{2(1-\rho_o)}{n(1-\mu\rho_o)}\sigma_0^2 \tag{3.2a}$$

If ρ_o is positive then the best value of μ which minimizes $V(\hat{\Omega})$ is zero. This suggests complete matching of units on the two occasions for estimating change.

If estimate of change is based on simple average values on both occasions then the estimator is given by

$$\hat{\Omega}' = \lambda(\hat{\bar{y}}' - \hat{\bar{x}}') + \mu(\hat{\bar{y}}'' - \hat{\bar{x}}'') \tag{3.3}$$

and the variance by

$$\frac{1}{n}[\sigma_{01}^2 + \sigma_{02}^2 - 2\lambda\rho_0\sigma_{01}\sigma_{02}] \tag{3.4}$$

A simplified expression for the variance is given by

$$\frac{2}{n}\sigma_0^2(1-\lambda\rho_o) \text{ if } \sigma_{01}^2 = \sigma_{02}^2 = \sigma_0^2 \tag{3.4a}$$

The percent gain in precision of $\hat{\Omega}$ over $\hat{\Omega}'$ (using 3.2a and 3.4a) is

$$\frac{\lambda\mu\rho_0^2}{1-\rho_0} \times 100 \tag{3.5}$$

4. Estimation of Sum Over Two Occasions

The best estimator of sum over two occasions, $\Sigma = \bar{\omega} + \bar{\eta}$ is given by $\hat{\Sigma} = \hat{M}_1 + \hat{M}_2$

$$= \frac{1}{1-\mu^2\rho_0^2}\left[\begin{array}{l}\left\{\lambda\mu\rho_0\frac{\sigma_{02}}{\sigma_{01}}(\hat{\bar{x}}'' - \hat{\bar{x}}') + \mu(1-\mu\rho_o^2)\hat{\bar{y}}'' + \lambda\hat{\bar{y}}'\right\} \\ + \left\{\mu\lambda\rho_0\frac{\sigma_{01}}{\sigma_{02}}(\hat{\bar{y}}'' - \hat{\bar{y}}') + \mu(1-\mu\rho_o^2)\hat{\bar{x}}'' + \lambda\hat{\bar{x}}'\right\}\end{array}\right] \tag{4.1}$$

Since $\hat{\bar{y}}'' + \hat{\bar{x}}''$ is an unbiased estimator of Σ, we have

$$V(\hat{\Sigma}) = \text{Cov}(\hat{\bar{y}}'' + \hat{\bar{x}}'', \hat{\Sigma})$$

$$= \frac{1}{n(1-\mu^2\rho_0^2)}[(1-\mu\rho_0^2)(\sigma_{01}^2+\sigma_{02}^2)+2\lambda\rho_0\sigma_{01}\sigma_{02}] \quad (4.2)$$

$$= \frac{2(1+\rho_0)}{n(1+\mu\rho_0)}\sigma_0^2 \text{ if } \sigma_{01}^2 = \sigma_{02}^2 = \sigma_0^2 \quad (4.2a)$$

If ρ_0 is positive the best replacement policy is to have $\mu = 1$, *i.e.*, by taking a completely independent sample on the second occasion for estimating sum over two occasions.

An estimator of sum over two occasions based on simple average values on both the occasions is given by

$$\hat{\Sigma}' = \lambda(\hat{\bar{y}}' + \hat{\bar{x}}') + N(\hat{\bar{y}}'' + \hat{\bar{x}}'') \quad (4.3)$$

with variance

$$V(\hat{\Sigma}') = \frac{1}{n}[\sigma_{01}^2+\sigma_{02}^2+2\lambda\rho_0\sigma_{01}\sigma_{02}] \quad (4.4)$$

$$= \frac{2}{n}(1+\lambda\rho_0)\sigma_0^2 \text{ if } \sigma_{01}^2 = \sigma_{02}^2 = \sigma_0^2 \quad (4.4a)$$

The percentage gain in precision of $\hat{\Sigma}$ over $\hat{\Sigma}'$ (using 4.2a and 4.4a) is given by

$$\frac{\lambda\mu\rho_0^2}{1+\rho_0} \times 100 \quad (4.5)$$

It may be seen that the coefficients in the estimators for current occasion, change and sum over two occasions involve unknown parametric values. Either their estimated values or known values from priori knowledge are taken. Some errors are likely to be there in the estimation and the optimality of the estimates will be affected.

5. *Gain in Precision of the Estimators* $\hat{M}_2$, $\hat{\Omega}$ *and* $\hat{\Sigma}$

To get an idea about gain in precision for the estimator for the current occasion and optimum values of μ and λ, we assume that the ratio of measurement error variance and the true variance are same *i.e.*, $\frac{\sigma_{\delta 1}^2}{\sigma_{t1}^2} = \frac{\sigma_{\delta 2}^2}{\sigma_{t2}^2} = \frac{\sigma_\delta^2}{\sigma_t^2}$. Values of ρ, ρ_e and different levels of measurement errors *i.e.*, $\frac{\sigma_\delta^2}{\sigma_t^2}$ were plugged in expressions (2.10) and (2.7) respectively. The results are presented in Table 1.

Table 1. Percent gain in precision of $\hat{M}_2$ over $\hat{M}'_2$

Levels of measurement errors	λ	k	% gain	λ	k	% gain
		$\rho = 0.4, \rho_e = 0.6$			$\rho = 0.6, \rho_e = 0.4$	
10%	0.475	0.418	4.82	0.448	0.582	10.31
20%	0.473	0.433	5.20	0.452	0.566	9.65
30%	0.472	0.466	5.59	0.455	0.554	9.05
		$\rho = 0.4, \rho_e = 0.8$			$\rho = 0.6, \rho_e = 0.8$	
10%	0.473	0.436	5.26	0.440	0.618	11.91
20%	0.469	0.466	6.16	0.436	0.633	12.73
30%	0.465	0.492	6.95	0.433	0.646	13.44

Similarly for the estimators of change and sum over two occasions the percent gain in precision was obtained for different values of ρ, ρ_e, $\frac{\sigma_\delta^2}{\sigma_t^2}$ and λ. The results are presented in Tables 2 and 3 respectively.

Table 2. Precent gain in precision of $\hat{\Omega}$ over $\hat{\Omega}'$

Levels of measurement errors	k	λ			k	λ		
		1/2	1/3	1/5		1/2	1/3	1/5
		% gain				% gain		
		$\rho = 0.4, \rho_e = 0.6$				$\rho = 0.6, \rho_e = 0.4$		
10%	0.418	7.47	6.64	4.78	0.582	20.27	18.02	12.97
20%	0.433	8.24	7.33	5.28	0.566	18.43	16.36	11.79
30%	0.446	8.98	7.98	5.75	0.554	17.21	15.28	11.01
		$\rho = 0.4, \rho_e = 0.8$				$\rho = 0.6, \rho_e = 0.8$		
10%	0.436	8.42	7.48	5.39	0.618	25.00	22.22	16.00
20%	0.466	10.15	9.03	6.50	0.633	27.24	24.20	17.44
30%	0.492	11.90	10.59	7.62	0.646	29.45	25.93	18.85

Table 3. Precent gain in precision of $\hat{\Sigma}$ over $\hat{\Sigma}'$

Levels of measurement errors	k	λ 1/2	λ 1/3	λ 1/4	k	λ 1/2	λ 1/3	λ 1/4
		% gain				% gain		
		$\rho = 0.4, \rho_e = 0.6$				$\rho = 0.6, \rho_e = 0.4$		
10%	0.418	3.08	2.73	2.31	0.582	5.35	4.75	4.01
20%	0.433	3.27	2.90	2.45	0.566	5.11	4.54	3.83
30%	0.446	3.43	3.05	2.58	0.554	4.94	4.38	3.70
		$\rho = 0.4, \rho_e = 0.8$				$\rho = 0.6, \rho_e = 0.8$		
10%	0.436	3.31	2.94	2.48	0.618	5.90	5.24	4.42
20%	0.466	3.70	3.29	2.78	0.633	6.13	5.44	4.60
30%	0.492	4.05	3.60	3.04	0.646	6.34	5.62	4.75

A close perusal of the foregoing results reveal that the gain in precision of estimator for the current occasion improves when the proportion of matched sample on the second occasion decreases while the precision of estimators for the change and sum over two occasions increases as the proportion of matched sample on the second occasion increases. Also the gain in precision increases with increase in correlation between true values and measurement errors on two occasions *i.e.*, ρ and ρ_e. It increases with increase in levels of measurement errors provided $\rho_e > \rho$ and decreases with increase in levels of measurement errors for $\rho > \rho_e$. It can also be seen that the gain in precision is modest for estimating sum over two occasions.

The gain in precision of these estimators vis-a-vis those free of measurement error is more whenever $\rho_e > \rho$ and vice-versa.

ACKNOWLEDGEMENTS

The authors are grateful to the referees for suggestions for improving the quality of the paper. The authors express sincere thanks to Shri D.P. Sharma, Technical Officer, for his help in the preparation of this paper.

REFERENCES

[1] Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A. and Sudman, S. (1991). *Measurement Errors in Surveys,* Wiley, New York..

[2] Patterson, H.D. (1950). Sampling on successive occasions with partial replacement of units. *Jour. Roy. Statist. Soc.,* **B, 12,** 241-255.

[3] Rao, C.R. (1952). Some theorems on MVUE, *Sankhya,* **12,** 27-42.

[4] Yates, F. (1949). *Sampling Methods for Censuses and Surveys.* Griffin and Co, London.