STATISTICS IN TRANSITION-new series, December 2010
Vol. 11, No. 3, pp. 555—562

ON EFFICIENT DIFFERENCE TYPE ESTIMATORS

A. K. P. C. Swain[1]

ABSTRACT

This paper considers a more efficient difference type estimator in a finite population set-up. Ratio type and regression type estimators are derived as special cases. Further efficiencies of these estimators are compared with classical ratio and regression estimators. Numerical illustrations are provided to compare efficiencies of different competitive estimators.

Key words: Difference type, Ratio type and Regression type estimators, Efficiencies of estimators, Auxiliary information, Simple random sampling without replacement.

1. Introduction

In sample surveys, ratio and regression estimators are often used to make use of auxiliary information to produce more efficient estimators compared to estimator based on simple mean per unit of the study variable.

The literature on ratio and regression methods is quite extensive. Many researchers, such as Bedi and Hajela (1984), Gupta (1978), Pandey (1980), Prasad (1986), Ray and Singh (1981), Rao (1978), Rao (1993), Ray and Sahai (1980), Srivastava (1967), Tailor and Sharma (2009) and others, have come forward to suggest more methods of improving precision of conventional ratio and regression estimators under certain conditions.

In the following we shall consider some more efficient difference type, ratio type and regression type estimators to estimate the population mean of the study variable in the presence of auxiliary information.

[1] Visiting Professor, Vivekananda Institute of Social Works and Social Sciences, Bhubaneswar-751001, India. E-mail: akpcs@rediffmail.com.

2. Generalized Difference type estimator

Let there be a finite population U consisting of N identifiable units $U_1, U_2, ..., U_N$. On the i^{th} unit a paired value (Y_i , X_i), $(i = 1, 2,..., \mathrm{N})$ is attached where Y_i and X_i $(i = 1, 2,..., \mathrm{N})$ are realized values of the study variable y and auxiliary variable x respectively.

Define $\overline{Y}$ and $\overline{X}$ as the population means of y and x respectively where $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ and $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$. For a simple random sample 's' of size n drawn without replacement from the finite population U, define the sample means of y and x as $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ respectively.

In order to estimate the population mean $\overline{Y}$, define the generalized difference type estimator

$$\hat{\overline{Y}}_d^* = \alpha\overline{y} + d\left(\overline{X} - \overline{x}\right), \tag{2.1}$$

where α is a suitably chosen scalar and d is a known constant or a random variable converging in probability to a constant.

$\hat{\overline{Y}}_d^*$ is a biased estimator with bias equal to

$$Bias\left(\hat{\overline{Y}}_d^*\right) = E\left(\hat{\overline{Y}}_d^*\right) - \overline{Y}$$

$$= \alpha\overline{Y} - \overline{Y} = (\alpha - 1)\overline{Y}, \tag{2.2}$$

The mean square error of $\hat{\overline{Y}}_d^*$ is given by

$$MSE\left(\hat{\overline{Y}}_d^*\right) = \alpha^2 V(\overline{y}) + d^2 V(\overline{x}) - 2\alpha d\ Cov(\overline{y}, \overline{x}) + (\alpha - 1)^2\overline{Y}^2 \tag{2.3}$$

The value of α which minimizes $MSE\left(\hat{\overline{Y}}_d^*\right)$ is

$$\alpha_{opt} = \frac{d\ Cov(\overline{y}, \overline{x}) + \overline{Y}^2}{V(\overline{y}) + \overline{Y}^2}, \tag{2.4}$$

Thus, the minimum MSE of $\hat{\overline{Y}}_d^*$ to $O\left(\frac{1}{n}\right)$ is given by

$$MSE\left(\hat{\hat{Y}}_d^*\right)_{\min} \cong \frac{\theta\left(S_y^2 + d^2 S_x^2 - 2dS_{xy}\right)}{\left(1+\theta C_y^2\right)^2}, \tag{2.5}$$

Where , $S_y^2 = \frac{1}{N-1}\sum\left(Y_i - \bar{Y}\right)^2$, $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$

$$S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right),\ C_y^2 = \frac{S_y^2}{\bar{Y}^2},$$

and $\theta = \left(\frac{1}{n} - \frac{1}{N}\right)$.

3. Ratio type estimator

Substituting $d = r = \frac{\bar{y}}{\bar{x}}$ in (2.1), the ratio type estimator is given by

$$\hat{\hat{Y}}_R^* = \alpha\bar{y} + r(\bar{X} - \bar{x}). \tag{3.1}$$

Assuming that the bias in r is negligible for large samples, the values of α, which minimizes $MSE\left(\hat{\hat{Y}}_R^*\right)$ is

$$\alpha_{opt} = \frac{1+\theta\rho\, C_y C_x}{1+\theta\, C_y^2}, \tag{3.2}$$

where C_y and C_x are the coefficients of variation of y and x respectively and ρ is the correlation coefficient between y and x.

Thus, the minimum MSE of $\hat{\hat{Y}}_R^*$ to $O\left(\frac{1}{n}\right)$ is,

$$MSE\left(\hat{\hat{Y}}_R^*\right) = \frac{\bar{Y}^2\left[\theta\left(C_y^2 + C_x^2 - 2\rho C_y C_x\right)\right]}{\left(1+\theta C_y^2\right)^2} \tag{3.3}$$

The classical ratio estimator $\hat{\hat{Y}}_R = \left(\frac{\bar{y}}{\bar{x}}\right)\bar{X}$ has the mean square error to $O\left(\frac{1}{n}\right)$, given by

$$MSE\left(\hat{\hat{Y}}_R\right) = \theta \overline{Y}^2 \left(C_y^2 + C_x^2 - 2\rho C_y C_x\right). \tag{3.4}$$

It may be verified that $MSE\left(\hat{\hat{Y}}_R^*\right) \leq MSE\left(\hat{\hat{Y}}_R\right)$, showing thereby that $\hat{\hat{Y}}_R^*$ is more efficient than $\hat{\hat{Y}}_R$.

4. Regression type estimators

Putting $d = \hat{\beta}$, where $\hat{\beta}$ is all estimate of regression coefficient of y on x, we have the regression type estimator (Kaur, 1985), given by

$$\hat{\hat{Y}}_{\text{Reg}}^* = \alpha \overline{y} + \hat{\beta}\left(\overline{X} - \overline{x}\right). \tag{4.1}$$

Neglecting bias in $\hat{\beta}$ for large sample, we have

$$Bias\left(\hat{\hat{Y}}_{\text{Reg}}^*\right) = \left(\alpha - 1\right)\overline{Y}. \tag{4.2}$$

The value of α which minimizes the $MSE\left(\hat{\hat{Y}}_{\text{Reg}}^*\right)$ is

$$\alpha_{opt} = \frac{1 + \theta\rho^2 C_y^2}{1 + \theta C_y^2}. \tag{4.3}$$

Thus, the minimum MSE of $\hat{\hat{Y}}_{\text{Reg}}^*$ is

$$MSE\left(\hat{\hat{Y}}_{\text{Reg}}^*\right) = \theta \overline{Y}^2 C_y^2 \left(1 - \rho^2\right)\left(1 + \theta C_y^2\right)^{-2}, \tag{4.4}$$

which is less than the approximate MSE of the conventional linear regression estimator $\hat{\hat{Y}}_{\text{Reg}} = \overline{y} + \hat{\beta}\left(\overline{X} - \overline{x}\right)$, given by

$$MSE\left(\hat{\hat{Y}}_{\text{Reg}}\right) = \theta \overline{Y}^2 C_y^2 \left(1 - \rho^2\right). \tag{4.5}$$

However, the estimator $\hat{\hat{Y}}_{\text{Reg}}^*$ happens to be a special case of a generalized estimator

$$\hat{\hat{Y}}_G^* = K_1 \overline{y} + K_2 \left(\overline{X} - \overline{x}\right), \tag{4.6}$$

where K_1 and K_2 are suitably chosen constants to be determined by minimizing $MSE\left(\hat{\bar{Y}}_G^*\right)$.

Thus,

$$MSE\left(\hat{\bar{Y}}_G^*\right) = K_1^2\ V(\bar{y}) - 2K_1K_2\ Cov\left(\bar{y}\ ,\ \bar{x}\right) + K_2^2\ V(\bar{x}) + \left(K_1 - 1\right)^2 \bar{Y}^2 . \quad (4.7)$$

The optimum values of K_1 and K_2 are

$$K_2 = K_1\beta \quad \text{and} \quad K_1 = \left[1 + \theta C_y^2\left(1-\rho^2\right)\right]^{-1}, \qquad (4.8)$$

where β is the population regression coefficient of y on x. As such, the minimum mean square error of $\hat{\bar{Y}}_G^*$ to $O\left(\frac{1}{n}\right)$ is given by

$$MSE\left(\hat{\bar{Y}}_G^*\right)_{\min} = \theta\bar{Y}^2 C_y^2\left(1-\rho^2\right)\left[1 + \theta C_y^2\left(1-\rho^2\right)\right]^{-2}, \qquad (4.9)$$

which is less than the approximate mean square error of $\hat{\bar{Y}}_{\mathrm{Re}g}$.

Bedi and Hajela (1984) proposed a modified regression estimator

$$\hat{\bar{Y}}_{BH} = w\left(\bar{y} + \beta\left(\bar{X} - \bar{x}\right)\right) \qquad (4.10)$$

where w is a suitably chosen constant so as to minimize the mean square error of $\hat{\bar{Y}}_{BH}$. It may be verified that $MSE\left(\hat{\bar{Y}}_{BH}\right)_{\min} = MSE\left(\hat{\bar{Y}}_G^*\right)$

Rao (1978) suggested a regression type estimator

$$\hat{\bar{Y}}_{\mathrm{R}ao}^* = K_1\bar{y} + K_2\bar{x}\,, \qquad (4.11)$$

where K_1 and K_2 are constants, chosen optimally subject to condition of unbiasedness i.e. $E\left(\hat{\bar{Y}}_{\mathrm{R}ao}^*\right) = \bar{Y}$, which implies $\left(K_1 - 1\right)\bar{Y} + K_2\bar{X} = 0$.

Minimizing $MSE\left(\hat{\bar{Y}}_{\mathrm{R}ao}^*\right)$ subject to condition for unbiasedness gives

$$K_{1(opt)} = \frac{C_x^2 - \rho C_y C_x}{C_y^2 + C_x^2 - 2\rho C_y C_x} \quad \text{and} \quad K_{2(opt)} = R \cdot \frac{C_y^2 - \rho C_y C_x}{C_y^2 + C_x^2 - 2\rho C_y C_x}, \quad (4.12)$$

where $R = \frac{\bar{Y}}{\bar{X}}$.

Substituting the optimum values of K_1 and K_2 in the expression for mean square error of $\hat{\bar{Y}}^*_{Rao}$, we have $O\left(\frac{1}{n}\right)$,

$$MSE\left(\hat{\bar{Y}}^*_{Rao}\right) = \theta \cdot \bar{Y}^2 \cdot \frac{\left(1-\rho^2\right)C_y^2 C_x^2}{C_y^2 + C_x^2 - 2\rho C_y C_x}. \tag{4.13}$$

$\hat{\bar{Y}}^*_{Rao}$ is more efficient than the conventional linear regression estimator $\hat{\bar{Y}}_{Reg}$ if $\rho < \frac{1}{2}\frac{C_y}{C_x}$, and is more efficient than $\hat{\bar{Y}}^*_{Reg}$ if $\rho < \frac{1}{2}\left(\frac{C_y}{C_x} - \theta C_y C_x\right)$.

Further, $\hat{\bar{Y}}^*_{Rao}$ is more efficient than the more efficient regression type estimator $\hat{\bar{Y}}^*_G$, if $\rho + \theta C_y C_x \left(1-\rho^2\right) < \frac{1}{2}\frac{C_y}{C_x}$.

5. Numerical illustration

Consider the following hypothetical populations to illustrate the performance of different estimators considered above.

Population 1 (Kaur, 1985)

$S_y^2 = 620$, $S_x^2 = 7619$, $S_{xy} = 1453$

$\bar{Y} = 26.30$, $\bar{X} = 117.28$

$\rho = 0.6685$, $n = 100$

Population 2 (Menendez and Reyes, 1998)

$S_y^2 = 8.40$, $S_x^2 = 3.31$, $\rho = 0.20$

$\bar{Y} = 12.0$, $\bar{X} = 16.6$, $N = 5000$, $n = 100$

Table 1. Percent Relative Efficiencies

Estimators	Population – 1	Population - 2
	Relative Efficiency	Relative Efficiency
$\bar{y}$	100	100
$\hat{\bar{Y}}_R$	176.39	97.62
$\hat{\bar{Y}}_R^*$	179.56	97.72
$\hat{\bar{Y}}_{\text{Reg}}$	180.80	104.17
$\hat{\bar{Y}}_{\text{Reg}}^*$	184.05	104.29
$\hat{\bar{Y}}_G^*$	182.60	104.28
$\hat{\bar{Y}}_{Rao}^*$	165.87	518.21

6. Conclusion

1. Although theoretically subject to approximations, $\hat{\bar{Y}}_R^*$ and $\hat{\bar{Y}}_{\text{Reg}}^*$ are more efficient than $\hat{\bar{Y}}_R$ and $\hat{\bar{Y}}_{\text{Reg}}$ respectively, the increase in efficiency for large sample sizes is only marginal.
2. When comparing $\hat{\bar{Y}}_{\text{Reg}}^*$ with $\hat{\bar{Y}}_G^*$, it may be seen that $\hat{\bar{Y}}_{\text{Reg}}^*$ is always more efficient than $\hat{\bar{Y}}_G^*$.
3. Under certain conditions, $\hat{\bar{Y}}_{Rao}^*$ is more efficient than $\hat{\bar{Y}}_{\text{Reg}}$ and $\hat{\bar{Y}}_{\text{Reg}}^*$ and $\hat{\bar{Y}}_G^*$. The empirical studies through hypothetical populations show that $\hat{\bar{Y}}_{Rao}^*$ is inferior to all estimators under consideration for the population–1 and there is large gain in precision in case of $\hat{\bar{Y}}_{Rao}^*$ to other estimators considered for illustration for population-2.

REFERENCES

BEDI, P.K. and HAJELA, D. (1984). An estimator for mean utilizing known coefficient of variation and auxiliary variable, J. Stat. Res., 18, 29–33.

COCHRAN, W.G. (1977). Sampling Techniques, 3rd ed., Wiley.

GUPTA, P.C. (1978). On some quadratic and higher degree ratio and product estimators, J. Ind. Soc. Ag. Stat., 30, 71–80.

KAUR, P. (1985). An efficient regression type estimator in survey sampling, Biom. Journal, 27, 107–110.

MENENDEZ, E. and REYES, A (1998). On an efficient Regression Type estimator, Biom. J., 40, 1, 79–84.

PANDEY, G.S. (1980). Product – Cum – Power estimators, Cal. Stat. Assoc. Bull. 29, 103–108.

PRASAD, B. (1986). Some unbiased estimators Versus mean per unit and ratio estimators in finite population sample surveys, Comm. in Stat. Th. and Meth., 15, 3647–3657.

RAO, I.S. (1978). On a method of using auxiliary information, contributions to statistics, edited by Department of Statistics, Utkal University, Bhubaneswar (India), pp. 65–69.

RAO, T.J. (1993). Auxiliary information in sample surveys: unpublished manuscript.

RAY, S.K. and SAHAI, A. (1980). Efficient families of ratio and product type estimators, Biometrika, 67, 211–215.

RAY, S.K. and SINGH, R.K. (1981). Difference – Cum – ratio type estimators. J. Ind. Stat. Assoc., 19, 147–151.

SRIVASTAVA, S.K. (1967). An estimator using auxiliary information in sample surveys, Cal. Stat. Assoc. Bull., 16, 121–132.

TAILOR, R and SHARMA, B (2009). A modified Ratio-cum-Product Estimator of finite population mean using known coefficient of variation and coefficient of kurtosis, Statistics in Transition-New Series, Vol – 10, PP. 15–24.