STATISTICS IN TRANSITION-new series, July 2009
Vol. 10, No. 1, pp. 15—24

A MODIFIED RATIO-CUM-PRODUCT ESTIMATOR OF FINITE POPULATION MEAN USING KNOWN COEFFICIENT OF VARIATION AND COEFFICIENT OF KURTOSIS

Rajesh Tailor[1] and Balkishan Sharma

ABSTRACT

This paper proposes a ratio-cum-product estimator of finite population mean using information on coefficient of variation and coefficient of kurtosis of auxiliary variate. The bias and mean squared error of the proposed estimator are obtained. It has been shown that the proposed estimator is more efficient than the sample mean estimator, usual ratio and product estimators and estimators proposed by Upadhyaya and Singh (1999) under certain given conditions. An empirical study is also carried out to demonstrate the merits of the proposed estimator over other estimators.

Key words: Coefficient of Variation, Coefficient of Kurtosis, Finite Population Mean, Bias and Mean Squared Error.

1. Introduction

In sample surveys, auxiliary information is used at both selection as well as estimation stages to improve the efficiency of the estimators. When the correlation between study variate and auxiliary variate is positive (high), the ratio method of estimation is used for estimating the population mean. On the other hand, if the correlation is negative, the product method of estimation envisaged by Robson (1957) is used. Sisodia and Dwivedi (1981) and Pandey and Dubey (1988) used coefficient of variation along with the population mean of auxiliary variate. Use of coefficient of kurtosis of auxiliary variate has also been in practice for improving the efficiency of the estimators of finite population mean. Upadhyaya and Singh (1999) and Singh et al. (2004) utilized coefficient of kurtosis of auxiliary variate for estimating the finite population mean. Rao and

[1] E-mail: tailorraj@gmail.com

Mudholkar (1967) and Singh and Espejo (2003) motivated authors to propose a ratio-cum-product estimator of finite population mean.

Let $U = (U_1, U_2 ..., U_N)$ be the finite population of size N and y and x be the study and auxiliary variates respectively. A sample of size n is drawn using simple random sampling without replacement to estimate the population mean $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ of study variate y.

The classical ratio and product estimators for $\bar{Y}$ are respectively given by

$$\bar{y}_R = \bar{y}\left(\frac{\bar{X}}{\bar{x}}\right) \tag{1.1}$$

and

$$\bar{y}_P = \bar{y}\left(\frac{\bar{x}}{\bar{X}}\right) \tag{1.2}$$

where $\bar{x} = \sum_{i=1}^{n} x_i / n$ and $\bar{y} = \sum_{i=1}^{n} y_i / n$. Here it is assumed that $\bar{X} = \sum_{i=1}^{N} x_i / N$, population mean of auxiliary variate is known.

Sisodia and Dwivedi (1981) suggested a ratio estimator of population men using coefficient of variation C_x of auxiliary variate as

$$\hat{\bar{Y}}_{SD} = \bar{y}\left(\frac{\bar{X} + C_x}{\bar{x} + C_x}\right) \tag{1.3}$$

Utilizing the information on coefficient of variation C_x and coefficient of kurtosis $\beta_2(x)$ of the auxiliary variate x, Upadhyaya and Singh (1999) proposed the following ratio and product estimators respectively

$$\hat{\bar{Y}}_{UP1} = \bar{y}\left(\frac{\bar{X}C_x + \beta_2(x)}{\bar{x}C_x + \beta_2(x)}\right) \tag{1.4}$$

$$\hat{\bar{Y}}_{UP2} = \bar{y}\left(\frac{\bar{x}C_x + \beta_2(x)}{\bar{X}C_x + \beta_2(x)}\right) \tag{1.5}$$

Singh et al. (2004) defined a ratio estimator using coefficient of kurtosis $\beta_2(x)$ which is

$$\hat{\bar{Y}}_{S} = \bar{y}\left(\frac{\bar{X} + \beta_2(x)}{\bar{x} + \beta_2(x)}\right) \tag{1.6}$$

2. Proposed Ratio-Cum-Product Estimator

The proposed ratio-cum-product estimator of population mean $\bar{Y}$ is

$$\hat{\bar{Y}}_M = \bar{y}\left[\alpha\left(\frac{\bar{X}C_x + \beta_2(x)}{\bar{x}C_x + \beta_2(x)}\right) + (1-\alpha)\left(\frac{\bar{x}C_x + \beta_2(x)}{\bar{X}C_x + \beta_2(x)}\right)\right] \tag{2.1}$$

where α is a suitably chosen scalar. It is to be noted that for α =1 and α = 0, $\hat{\bar{Y}}_M$ reduces to the estimators $\hat{\bar{Y}}_{UP1}$ and $\hat{\bar{Y}}_{UP2}$ respectively suggested by Upadhyaya and Singh (1999). Thus $\hat{\bar{Y}}_{UP1}$ and $\hat{\bar{Y}}_{UP2}$ are particular case of proposed estimator $\hat{\bar{Y}}_M$.

To obtain the bias and MSE of $\hat{\bar{Y}}_M$, we write
$\bar{y} = \bar{Y}(1+e_0)$ and $\bar{x} = \bar{X}(1+e_1)$ such that $E(e_0) = E(e_1) = 0$ and

$$E(e_0^2) = \left(\frac{1}{n} - \frac{1}{N}\right)C_y^2, \qquad E(e_1^2) = \left(\frac{1}{n} - \frac{1}{N}\right)C_x^2, \qquad E(e_0 e_1) = \left(\frac{1}{n} - \frac{1}{N}\right)\rho\, C_y\, C_x$$

where

$$C_y = S_y / \bar{Y}, \qquad C_x = S_x / \bar{X}, \qquad \rho = S_{yx} / S_x S_y, \qquad K = \rho C_y / C_x,$$

$$S_x^2 = \sum_{i=1}^{N}(x_i - \bar{X})^2 /(N-1), \qquad S_y^2 = \sum_{i=1}^{N}(y_i - \bar{Y})^2 /(N-1),$$

$$S_{xy} = \sum_{i=1}^{N}(y_i - \bar{Y})(x_i - \bar{X})/(N-1).$$

Expressing (2.1) in terms of $e_i^{'}s$ we have (i=1,2)

$$\hat{\bar{Y}}_M = \bar{Y}(1+e_0)\left[\alpha(1+t_3\, e_1)^{-1} + (1-\alpha)(1+t_3\, e_1)\right], \tag{2.2}$$

where $t_3 = \bar{X}C_x \big/ (\bar{X}C_x + \beta_2(x))$.

To the first degree of approximation, the bias and mean squared error of $\hat{\bar{Y}}_M$ are respectively given as

$$B(\hat{\bar{Y}}_M) = \frac{(1-f)}{n}\bar{Y}t_3\, C_x^2\left[K + t_3(t_3 - 2K)\right], \tag{2.3}$$

and

$$MSE(\hat{\bar{Y}}_M) = \frac{(1-f)}{n}\bar{Y}^2\left[C_y^2 + (1-2\alpha)t_3\, C_x^2\{(1-2\alpha)t_3 + 2K\}\right]. \tag{2.4}$$

Thus the estimator $\hat{\bar{Y}}_M$ with $\alpha = K/(2K - t_3)$ is almost unbiased. It is also observed from (2.3) that the bias of $\hat{\bar{Y}}_M$ is negligible for large sample.

Mean squared error of $\hat{\bar{Y}}_M$ in (2.4) is minimized for

$$\alpha = \frac{(t_3 + K)}{2t_3} = \alpha_0 \,(\text{say}) \tag{2.5}$$

Substitution of (2.5) in (2.1) yields the asymptotically optimum estimator (AOE) for

$$\bar{Y} \text{ as } \hat{\bar{Y}}_M^{(opt)} = \frac{\bar{y}}{2t_3}\left[(t_3 + K)\left(\frac{\bar{X}C_x + \beta_2(x)}{\bar{x}C_x + \beta_2(x)}\right) + (t_3 - K)\left(\frac{\bar{x}C_x + \beta_2(x)}{\bar{X}C_x + \beta_2(x)}\right)\right] \tag{2.6}$$

Putting (2.5) in (2.3) and (2.4), we get the bias and variance of $\hat{\bar{Y}}_M^{(opt)}$ respectively as

$$B\left(\hat{\bar{Y}}_M^{(opt)}\right) = \frac{(1-f)}{2n}\bar{Y}\, C_x^2 (t_3 - K)(t_3 + 2K), \tag{2.7}$$

and

$$MSE\left(\hat{\bar{Y}}_M^{(opt)}\right) = \frac{(1-f)}{n} S_y^2 (1 - \rho^2). \tag{2.8}$$

It is clear that mean squared error of $\hat{\bar{Y}}_M^{(opt)}$ is same as that of the approximate variance of the usual linear regression estimator $\bar{y}_{lr} = \bar{y} + \hat{\beta}(\bar{X} - \bar{x})$, where $\hat{\beta}$ is the sample regression coefficient of y on x.

3. Efficiency Comparisons

Under simple random sampling without replacement (SRSWOR), variance of sample mean $\bar{y}$ is

$$V(\bar{y}) = \frac{(1-f)}{n}\bar{Y}^2 C_y^2, \tag{3.1}$$

and the variance of $\bar{y}_R$, $\bar{y}_P$, $\hat{\bar{Y}}_{SD}$, $\hat{\bar{Y}}_S$, $\hat{\bar{Y}}_{UP1}$ and $\hat{\bar{Y}}_{UP2}$ to the first degree of approximation are respectively given by

$$MSE(\bar{y}_R) = \frac{(1-f)}{n}\bar{Y}^2\left[C_y^2 + C_x^2(1-2K)\right] \tag{3.2}$$

$$MSE(\bar{y}_P) = \frac{(1-f)}{n}\bar{Y}^2\left[C_y^2 + C_x^2(1+2K)\right] \tag{3.3}$$

$$MSE(\hat{\bar{Y}}_{SD}) = \frac{(1-f)}{n}\bar{Y}^2\left[C_y^2 + t_1\, C_x^2(t_1 - 2K)\right] \tag{3.4}$$

$$MSE(\hat{\bar{Y}}_{S}) = \frac{(1-f)}{n}\bar{Y}^2\left[C_y^2 + t_2\, C_x^2(t_2 - 2K)\right] \tag{3.5}$$

$$MSE(\hat{\bar{Y}}_{UP2}) = \frac{(1-f)}{n}\bar{Y}^2\left[C_y^2 + t_3\, C_x^2(t_3 - 2K)\right] \tag{3.6}$$

$$MSE(\hat{\bar{Y}}_{UP2}) = \frac{(1-f)}{n}\bar{Y}^2\left[C_y^2 + t_3\, C_x^2(t_3 + 2K)\right] \tag{3.7}$$

From (2.4) and (3.1), it is observed that $\hat{\bar{Y}}_M$ is more efficient than usual unbiased estimator $\bar{y}$ if

$$\left.\begin{array}{ll} either & \frac{1}{2} < \alpha < \left(\frac{1}{2} + \frac{K}{t_3}\right) \\ or & \left(\frac{1}{2} + \frac{K}{t_3}\right) < \alpha < \frac{1}{2} \end{array}\right\} \tag{3.8}$$

Comparison of (2.4) and (3.2) shows that $\hat{\bar{Y}}_M$ is more efficient than usual ratio estimator

$$\bar{y}_R \quad \left.\begin{array}{ll} either & \frac{(1+t_3)}{2t_3} < \alpha < \left(\frac{t_3 + 2K - 1}{2t_3}\right) \\ or & \left(\frac{t_3 + 2K - 1}{2t_3}\right) < \alpha < \frac{(1+t_3)}{2t_3} \end{array}\right\} \tag{3.8}$$

From (2.4) and (3.3) it is clear that $\hat{\bar{Y}}_M$ would be more efficient than $\bar{y}_P$ if

$$\left.\begin{array}{ll} either & \left(\dfrac{t_3+2K+1}{2t_3}\right) < \alpha < \dfrac{(t_3-1)}{2t_3} \\ or & \dfrac{(t_3-1)}{2t_3} < \alpha < \end{array}\right\} \tag{3.9}$$

Comparing (2.4) and (3.4), it is observed that $\hat{\bar{Y}}_M$ is more efficient than Sisodia and Dwivedi (1981) estimator $\hat{\bar{Y}}_{SD}$ if

$$\left.\begin{array}{ll} either & 1 < \alpha < \dfrac{K}{t_1} \\ or & \dfrac{K}{t_1} < \alpha < 1 \end{array}\right\} \tag{3.10}$$

Comparing (2.4) and (3.5), it is observed that $\hat{\bar{Y}}_M$ is more efficient than Singh et al. (2004) $\hat{\bar{Y}}_S$ if

$$\left.\begin{array}{ll} either & 1 < \alpha < \dfrac{K}{t_2} \\ or & \dfrac{K}{t_2} < \alpha < 1 \end{array}\right\} \tag{3.11}$$

Comparing (2.4) and (3.6), it is observed that $\hat{\bar{Y}}_M$ is more efficient than Upadhyaya and Singh (1999) estimator $\hat{\bar{Y}}_{UP1}$

$$\left.\begin{array}{ll} either & 1 < \alpha < \dfrac{K}{t_3} \\ or & \dfrac{K}{t_3} < \alpha < 1 \end{array}\right\} \tag{3.12}$$

Further, from (2.4) and (3.5), it is seen that $\hat{\bar{Y}}_M$ is more efficient than Upadhyaya and Singh (1999) estimator $\hat{\bar{Y}}_{UP2}$ if

$$\left.\begin{array}{ll} \textit{either} & 0<\alpha<\left(1+\dfrac{K}{t_3}\right) \\ \textit{or} & \left(1+\dfrac{K}{t_3}\right)<\alpha<0 \end{array}\right\} \quad (3.13)$$

4. Empirical Study

To analyze the performance of the proposed estimator in comparison to other estimators, four natural population data sets are being considered. The descriptions of the populations are given below.

Population-I [Source: Das (1988)]

The population consists of 278 village town/wards under Gajole police station of Malada district of West Bengal, India (in fact, only the villages of town/wards which are shown as inhabited and common in both census 1961 and census 1971 lists have been considered). The variates considered are

x: The number of agricultural labourers for 1961

y: The number of agricultural labourers for 1971

$\bar{Y}$=39.0680, $\bar{X}$=25.1110, C_y=1.4451,

C_x=1.6198, ρ=0.7213, $\beta_2(x)$=38.8898

Population-II [Source: Das (1988)]

It consists of 142 cities of India with population (number of persons) 100,000 and above; the character x and y being

x: Census population in the year 1961

y: Census population in the year 1971

$\bar{Y}$=4015.2183, $\bar{X}$=2900.3872, C_y=2.1118,

C_x=2.1971, ρ=.9948, $\beta_2(x)$=48.1567

Population-III [Source: Cochran (1977)]

The variates are defined as follows:

y: The number of persons per block

x: The number of rooms per block

$\bar{Y}$=101.1, $\bar{X}$=58.80, C_y=0.14450,

C_x=0.1281 , ρ=0.6500, $\beta_2(x)$=2.2387.

Population-IV [Source: Pandey and Dube (1988)]

The population constants are as follows:

N=20, n=8,

$\overline{Y} = 19.55$ $\overline{X} = 18.8$, $C_y^2 = 0.1262$,

$C_x^2 = 0.1555$, $\rho_{yx} = -0.9199$, $\beta_2(x) = 3.0613$

Table 4.1. Range of α in which $\hat{\hat{Y}}_M$ is better than $\bar{y}$, $\bar{y}_R$, $\bar{y}_p$ $\hat{\hat{Y}}_{SD}$, $\hat{\hat{Y}}_S$, $\hat{\hat{Y}}_{UP1}$ and $\hat{\hat{Y}}_{UP2}$

Popu-lation	Range of α in which is $\hat{\hat{Y}}_M$ better than							Opti-mum value of α
	$\bar{y}$	$\bar{y}_R$	$\bar{y}_P$	$\hat{\hat{Y}}_{SD}$	$\hat{\hat{Y}}_S$	$\hat{\hat{Y}}_{UP1}$	$\hat{\hat{Y}}_{UP2}$	α_0
I	(0.5, 1.463)	(1.004, .989)	(-0.04, .1.967)	(1.00, .957)	(1.00, .972)	(.963, 1.00)	(0.0,1.963)	(0.982)
II	(0.5, 1.759)	(0.989, 0.787)	(-0.478,2.737)	(1.00, .685)	(1.00,1.64)	(1, 1.258)	(0.0, 2.59)	(1.129)
III	(0.5, 1.451)	(0.683, 0.803)	(-0.149, 2.10)	(1.00, .735)	(1.00, .761)	(0.951,1.00)	(0.0,1.951)	(0.976)
IV	(0.5, -.670)	(0.683, -1.378)	(-.064.8, .035)	(1.00, -.846)	(1.00, -.9632)	(1.00, -1.17)	(0.0,-.1709)	(-.0857)

Table 4.2. Percent relative efficiencies of $\bar{y}$, $\bar{y}_R$, $\bar{y}_p$ $\hat{\hat{Y}}_{SD}$, $\hat{\hat{Y}}_S$, $\hat{\hat{Y}}_{UP1}$, $\hat{\hat{Y}}_{UP2}$ and $\hat{\hat{Y}}_M$ or $\hat{\hat{Y}}_M^{(opt)}$ with respect to $\bar{y}$

Estimators	Population I	Population II	Population III	Population IV
$\bar{y}$	100.00	100.00	100.00	100.00
$\bar{y}_R$	156.40	8031.10	157.87	23.39
$\bar{y}_p$	*	*	*	526.45
$\hat{\hat{Y}}_{SD}$	169.57	8077.28	158.10	23.91
$\hat{\hat{Y}}_S$	178.90	8935.82	161.52	27.27
$\hat{\hat{Y}}_{UP1}$	199.32	8473.89	172.83	32.65
$\hat{\hat{Y}}_{UP2}$	*	*	*	582.03
$\hat{\hat{Y}}_M$ ($\hat{\hat{Y}}_M^{(opt)}$)	208.45	9640.45	173.16	650.26

* *Data are not applicable.*

Table 4.2 shows that there is a significant gain in efficiency by using proposed estimator $\hat{\bar{Y}}_M$ over unbiased estimator $\bar{y}$, usual ratio estimator $\bar{y}_R$, product estimator $\bar{y}_p$, Sisodiya and Dwivedi (1981) estimator $\hat{\bar{Y}}_{SD}$, Singh et al. (2004) estimator $\hat{\bar{Y}}_S$ and Upadhyay and Singh (1999) estimators $\hat{\bar{Y}}_{UP1}$ and $\hat{\bar{Y}}_{UP2}$.

Table 4.1 provides the wide range of α in which suggested estimator $\hat{\bar{Y}}_M$ ($\hat{\bar{Y}}_M^{(opt)}$) is more efficient then all estimators considered in this paper. This shows that even if the scalar α deviates from its optimum value (α_{opt}), the suggested estimator $\hat{\bar{Y}}_M^{(opt)}$ will yield better estimates than $\bar{y}$, $\bar{y}_R$, $\bar{y}_p$, $\hat{\bar{Y}}_{SD}$, $\hat{\bar{Y}}_S$, $\hat{\bar{Y}}_{UP1}$ and $\hat{\bar{Y}}_{UP2}$. Therefore, suggested estimator $\hat{\bar{Y}}_M$ is recommended for use in practice.

REFERENCES

COCHRAN, W.G. (1977). Sampling Techniques. Third U. S. Edition. Wiley Eastern Limited. 325.

DAS, A.K. (1988). Contribution to the theory of sampling strategies based on auxiliary information. Ph.D. Thesis, BCKV, West Bengal, India.

PANDEY, B.N. and DUBEY, V. (1988). Modified product estimator using coefficient of variation of auxiliary variable. Assam Stat. Rev., 2, 64–66.

RAO, P.S. R.S. and MUDHOLKAR, G.S. (1967). Generalised multivariate estimators for the mean of a finite population. Jour. Ind. Stat. Assoc., 62, 1009–1012.

ROBSON, D.S. (1957). Application of multivariate polykays to the theory of unbiased ratio-type estimation. Jour. Amer. Stat. Assoc., 59, 1225–1226.

SINGH, H.P., and ESPEJO, M.R. (2003). On linear regression and ratio estimator using coefficient of variation of auxiliary variate. Statistician, 52,1, 59–67.

SINGH, H.P. TAILOR, R. (2005). Estimation of finite population mean using known correlation coefficient between auxiliary characters. Statistica, Anno LXV, n.4, pp. 407–418.

SINGH, H.P., KAKRAN, TAILOR, R. and TAILOR, R. (2004). A modified ratio estimator of population mean. Jour. Ind. Soc. Agri. Stat., 33(1), 13–18.

SISODIA, B.V.S. and DWIVEDI, V.K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. Jour. Ind. Soc. Agri. Stat., 33(1), 13–18.

UPADHYAYA, L.N. and SINGH, H.P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. Biometrical Jour., 41(5), 627–636.