Songklanakarin J. Sci. Technol.
41 (1), 117-122, Jan. – Feb. 2019

http://www.sjst.psu.ac.th

Original Article

Ratio estimators of population means using quartile function of auxiliary variable using double sampling

Nuanpan Lawson*

Department of Applied Statistics, Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok, Bang Sue, Bangkok, 10800 Thailand

Received: 9 August 2016; Revised: 28 September 2017; Accepted: 5 October 2017

Abstract

Ratio estimators have been used in survey sampling for estimating population mean when the population mean of an auxiliary variable is known. However, as the auxiliary information is not always available we cannot gain any benefit from using ratio estimators to increase the efficiency of the population mean estimator. New ratio estimators for population mean obtained using the quartile function of an auxiliary variable using double sampling have been proposed based upon the modified ratio estimators proposed by Subramani and Kumarapandiyan (2012). We propose the use of double sampling in order to estimate the unknown quartile function of an auxiliary variable. A simulation study has been conducted to compare these new ratio estimators with the classical ratio estimator using a percent relative efficiency using double sampling. The results show that the proposed estimators perform better than the classical ratio estimator using double sampling.

Keywords: ratio estimators, auxiliary variable, quartile function, double sampling, percent relative efficiency

1. Introduction

Many estimation techniques that require advanced knowledge of known auxiliary variables have been used in survey sampling. Estimating the population mean of study variable Y using ratio estimators is one of the most well-known techniques that requires knowledge of auxiliary information when auxiliary variable X and study variable Y are highly positively correlated. Ratio estimators are broadly used in social study, business, econometric, government and research study where it is important to make more precise estimators for a variable of interest. For example; the government need to address the issue of over unemployment, an unemployment rate can be estimated using the ratio of unemployment people and labour force size. Cochran (1977) proposed utilisation of the relationship between the variable of interest and auxiliary variable to improve estimations made of the population mean estimator for the variable of interest. The classical ratio estimator is given as below.

$$\hat{\bar{Y}}_R = \frac{\bar{y}}{\bar{x}}\bar{X}, \tag{1}$$

where $\bar{y}$ and $\bar{x}$ are the sample mean of study variable Y and auxiliary variable X respectively and $\bar{X}$ is the population mean of auxiliary variable X.

Later, several authors used ratio estimators to estimate population mean with the use of known population values of auxiliary variables such as coefficients of variation, kurtosis, skewness, correlation coefficient and median (Pandey & Dubey, 1988; Singh & Tailor, 2003, 2005; Singh & Upadhyaya, 1986; Sisodia & Dwivedi, 1981; Yan & Tian, 2010). Subramani and Kumarapandiyan (2012) proposed adjusting the classical ratio estimator by utilising the known quartile function of auxiliary variables; the first quartile, the third quartile, inter-quartile range, semi-quartile range and quartile average in simple random sampling, which were more efficient than some existing ratio estimators. The adjusted ratio estimator(s) proposed by Subramani and Kumarapandiyan (2012) is as follows:

*Corresponding author
Email address: nuanpan.n@sci.kmutnb.ac.th

$$\hat{\bar{Y}}_{SK1}=\bar{y}\left(\frac{\bar{X}+Q_3}{\bar{x}+Q_3}\right),\quad \hat{\bar{Y}}_{SK2}=\bar{y}\left(\frac{\bar{X}+Q_r}{\bar{x}+Q_r}\right),\quad \hat{\bar{Y}}_{SK3}=\bar{y}\left(\frac{\bar{X}+Q_d}{\bar{x}+Q_d}\right),\quad \hat{\bar{Y}}_{SK4}=\bar{y}\left(\frac{\bar{X}+Q_a}{\bar{x}+Q_a}\right), \tag{2}$$

where Q_1 and Q_3 denote the first and the third quartiles of auxiliary variable X respectively. Q_r denotes the inter-quartile range of auxiliary variable X, $Q_r = Q_3 - Q_1$, Q_d denotes the semi-quartile range of auxiliary variable X, $Q_d = \frac{(Q_3 - Q_1)}{2}$ and Q_a denotes the quartile average of auxiliary variable X, $Q_a = \frac{(Q_3 + Q_1)}{2}$.

However, if we lack information on the population mean of auxiliary variable X, a sampling technique called double sampling or two phase sampling proposed by Neyman (1938) can be used in order to estimate the unknown value for the population mean of an auxiliary variable. Let N be the number of units in a population U = $U_1,\ldots,U_N$. A large sample of n' units is selected with simple random sampling without replacement (SRSWOR) in the first phase sampling in order to collect the information from an auxiliary variable because it is quick and cheap to obtain when compared to a study variable. Then a smaller sample size n ($n < n'$) is selected in the second phase using SRSWOR in order to obtain the information on both the study variable and the auxiliary variable. Let $\bar{y} = \sum_{i=1}^{n} y_i / n$ and $\bar{x} = \sum_{i=1}^{n} x_i / n$ be the sample means of the study variable and the auxiliary variable from the second phase sampling based on subsample size n. If the population mean $\bar{X}$ is unknown, the classical ratio estimator in (1) using double sampling is defined as follows.

$$\hat{\bar{Y}}_{DR} = \frac{\bar{y}}{\bar{x}}\bar{x}', \tag{3}$$

where $\bar{x}' = \sum_{i=1}^{n'} x_i / n'$ is an unbiased estimator of population mean $\bar{X}$ from the first phase sampling based on sample size n'.

The ratio estimator is a biased estimator but this bias will be smaller as the sample gets larger. The Jackknife method was suggested by Quenouille (1956) reducing the bias of the estimator further.

As we mentioned earlier, various authors have suggested using some auxiliary information alongside the population mean including the quartile functions, which usually are not available. In this paper, we propose to extend the estimators proposed by Subramani and Kumarapandiyan using double sampling. We propose to estimate the unknown function of the quartiles of an auxiliary variable using double sampling. Moreover, we consider the Jackknife variance estimator in estimating the variance for combined estimators. The proposed estimators will be compared with the classical ratio estimator using double sampling via a simulation study.

2. Materials and Methods

We assume that the population mean and quartile functions of an auxiliary variable used in Subramani and Kumarapandiyan's (2012) estimator in simple random sampling are unknown and we propose to estimate these population values using double sampling. We propose to replace unknown values of the population mean $\bar{X}$ and quartile function of auxiliary variables Q_3, Q_r, Q_d and Q_a in (2) with estimated values of the sample mean $\bar{x}'$ and sample of the function of quartiles of an auxiliary variable Q_3', Q_r', Q_d' and Q_a' respectively using SRSWOR in the first phase sampling based on sample size n' and $\bar{x}$ and $\bar{y}$ in (2) are calculated in the second phase sampling using SRSWOR based on sample size n ($n < n'$). We called the new four proposed estimators $\hat{\hat{Y}}_{DR_1}$, $\hat{\hat{Y}}_{DR_2}$, $\hat{\hat{Y}}_{DR_3}$ and $\hat{\hat{Y}}_{DR_4}$ respectively.

We also suggest combined ratio estimators designed to minimize the variance in population mean of the variable of interest. The Jackknife variance estimator has been considered and extended from Rao and Sitter (1995). We combined two selected estimators ($\hat{\hat{Y}}_{DR_1}$ and $\hat{\hat{Y}}_{DR_2}$) which gave us a smaller variance than the other estimators using the Jackknife method. The combined ratio estimator $\hat{\hat{Y}}_{RC}$ is given as follows

$$\hat{\hat{Y}}_{RC} = \omega\hat{\hat{Y}}_{DR_1} + (1-\omega)\hat{\hat{Y}}_{DR_2} \tag{4}$$

The variance is given below.

$$Var(\hat{\hat{Y}}_{RC}) = \omega^2 Var(\hat{\hat{Y}}_{DR_1}) + (1-\omega)^2 Var(\hat{\hat{Y}}_{DR_2}) + 2\omega(1-\omega)COV(\hat{\hat{Y}}_{DR_1}, \hat{\hat{Y}}_{DR_2}). \tag{5}$$

The objective of this estimator is to find an estimator that minimises the variance of a population mean estimator of the variable of interest. We find the first derivative in respect to ω and set it equal to zero then ω is obtained as follows.

$$\omega = \frac{Var(\hat{\hat{Y}}_{DR_2})-COV(\hat{\hat{Y}}_{DR_1},\hat{\hat{Y}}_{DR_2})}{Var(\hat{\hat{Y}}_{DR_1})+Var(\hat{\hat{Y}}_{DR_2})-2COV(\hat{\hat{Y}}_{DR_1},\hat{\hat{Y}}_{DR_2})}, \tag{6}$$

where ω is the weighting from the combined ratio estimator that minimized the variance in the population mean estimator. $Var(\hat{\hat{Y}}_{DR_1})$ and $Var(\hat{\hat{Y}}_{DR_2})$ are the variance of the proposed estimators adjusting by using Q_3 and Q_r respectively. $COV(\hat{\hat{Y}}_{DR_1},\hat{\hat{Y}}_{DR_2})$ is the covariance between estimator $\hat{\hat{Y}}_{DR_1}$ and $\hat{\hat{Y}}_{DR_2}$. We extended the Jackknife variance estimator from Rao and Sitter using double sampling. The adjusted Jackknife variance estimator is shown as follows:

$$v(\hat{\hat{Y}}_{DR_j})=\frac{n-1}{n}\sum_{j=1}^{n}(\hat{\hat{Y}}_{DR(j)}-\hat{\hat{Y}}_{DRn})^2;\ j=1,2,...,n, \quad (7)$$

and

$$COV(\hat{\hat{Y}}_{DR_{1j}},\hat{\hat{Y}}_{DR_{2j}})=\frac{n-1}{n}\sum_{j=1}^{n}(\hat{\hat{Y}}_{R1(j)}-\hat{\hat{Y}}_{R1n})(\hat{\hat{Y}}_{R2(j)}-\hat{\hat{Y}}_{R2n});\quad j=1,2,...,n, \quad (8)$$

where $\hat{\hat{Y}}_{Rin}=\frac{1}{n}\sum_{j=1}^{n}\hat{\hat{Y}}_{R(j)}$, i = 1, 2. $\bar{x}'$ is sample mean from the first phase of sampling and $\bar{x}$ and $\bar{y}$ are the sample mean of the auxiliary and the variable of interest that have been estimated from the second phase sampling respectively, where $\bar{y}_{(j)}=\frac{n\bar{y}-y_j}{n-1}$, $\bar{x}_{(j)}=\frac{n\bar{x}-x_j}{n-1}$, $\bar{x}'_{(j)}=\frac{n'\bar{x}'-x_j}{n'-1}$.

3. Results and Discussion

We generate the population size N=1,000. A sample of n' = 200 units is selected in the first phase sampling then a sample size n (n = 20, 60 and 100) is selected in the second phase sampling with simple random sampling without replacement. The correlation between variable of interest and auxiliary variable (ρ_{xy}) is 0.5 and 0.8. The value of the variables of interest (X,Y) are generated from the bivariate normal distribution with different means and variances as shown in Tables 1 and 2. In Table 1 the means of variables X and Y are equal to 2 and 4 respectively and the variance for both variables is equal to 1. In Table 2 the means of variables X and Y are equal to 200 and 400 respectively and the variance for variable X is equal to 100, for Y it is equal to 9. A percentage relative efficiency (PRE) has been used to compare the performance of the proposed estimators with the classical ratio estimator using double sampling. The PRE of an estimator with respect to the classical ratio estimator is defined by $PRE(.,\hat{\hat{Y}}_{DR})=MSE(\hat{\hat{Y}}_{DR})/MSE(.)\times 100$. The simulation is repeated 10,000 times. The results are presented in Tables 1 and 2.

Table 1. Mean square error and percent relative efficiency of the proposed estimators over the classical ratio estimator under double sampling for N = 1000, n' = 200, $\bar{X}$ = 2, $\bar{Y}$ = 4 and $V(\bar{X})=V(\bar{Y})$ = 1.

ρ_{xy}	Estimator	n=20		n=60		n=100	
		MSE	PRE	MSE	PRE	MSE	PRE
0.5	$\hat{\hat{Y}}_{DR}$	0.1456	100.00	0.0390	100.00	0.0186	100.00
	$\hat{\hat{Y}}_{DR_1}$	0.0430	338.80	0.0141	276.42	0.0082	226.91
	$\hat{\hat{Y}}_{DR_2}$	0.0589	247.03	0.0181	214.45	0.0099	187.73
	$\hat{\hat{Y}}_{DR_3}$	0.0827	176.20	0.0240	161.98	0.0123	150.21
	$\hat{\hat{Y}}_{DR_4}$	0.0485	300.14	0.0155	250.93	0.0088	221.30
	$\hat{\hat{Y}}_{RC}$	**0.0408**	**356.30**	**0.0137**	**282.56**	**0.0081**	**228.65**
0.8	$\hat{\hat{Y}}_{DR}$	0.0894	100.00	0.0247	100.00	0.0127	100.00
	$\hat{\hat{Y}}_{DR_1}$	0.0202	443.28	0.0081	305.31	0.0057	223.03
	$\hat{\hat{Y}}_{DR_2}$	0.0272	328.90	0.0099	250.58	0.0064	196.91
	$\hat{\hat{Y}}_{DR_3}$	0.0424	210.88	0.0136	181.48	0.0080	157.89
	$\hat{\hat{Y}}_{DR_4}$	0.0217	411.66	0.0085	291.11	0.0059	216.50
	$\hat{\hat{Y}}_{RC}$	**0.0200**	**445.68**	**0.0080**	**305.90**	**0.0056**	**223.07**

From Tables 1 and 2 we can see similar results, we can see clearly that the proposed estimators perform a lot better than the classical ratio estimator using double sampling. The combined ratio estimator performs the best as it has the highest percentage relative efficiency when compared to the classical ratio estimator using double sampling. The percentage of relative efficiency is improved when the correlation between the variable of interest and the auxiliary variable increases and also when the sample size is increased.

Table 2. Mean square error and percent relative efficiency of the proposed estimators over the classical ratio estimator under double sampling for N = 1000, $n' = 200$, $\bar{X} = 200$, $\bar{Y} = 400$ and $V(\bar{X}) = 100$ and $V(\bar{Y}) = 9$.

ρ_{xy}	Estimator	n=20		n=60		n=100	
		MSE	PRE	MSE	PRE	MSE	PRE
0.5	$\hat{\bar{Y}}_{DR}$	15.0360	100.00	3.9513	100.00	1.6983	100.00
	$\hat{\bar{Y}}_{DR_1}$	3.3116	454.04	0.8925	442.69	0.3995	425.02
	$\hat{\bar{Y}}_{DR_2}$	13.0873	114.89	3.4532	114.42	1.4872	114.19
	$\hat{\bar{Y}}_{DR_3}$	14.0117	107.31	3.6898	107.09	1.5875	106.98
	$\hat{\bar{Y}}_{DR_4}$	3.4320	438.13	0.9233	427.96	0.4126	411.60
	$\hat{\bar{Y}}_{RC}$	**3.2886**	**457.22**	**0.8904**	**443.75**	**0.3987**	**425.86**
0.8	$\hat{\bar{Y}}_{DR}$	13.4840	100.00	3.5413	100.00	1.5260	100.00
	$\hat{\bar{Y}}_{DR_1}$	2.5471	529.40	0.6901	513.15	0.3143	485.48
	$\hat{\bar{Y}}_{DR_2}$	11.6332	115.91	3.0686	115.40	1.3257	115.12
	$\hat{\bar{Y}}_{DR_3}$	12.5110	107.78	3.2930	107.54	1.4208	107.41
	$\hat{\bar{Y}}_{DR_4}$	2.6545	507.98	0.7175	493.55	0.3260	468.16
	$\hat{\bar{Y}}_{RC}$	**2.5235**	**534.35**	**0.6878**	**514.87**	**0.3135**	**486.70**

In Figures 1, and 2 we can see that the proposed estimators perform better than the classical ratio estimator using double sampling. The bias gets smaller as the sample and the correlation between the variable of interest and the auxiliary variable gets larger. The combined ratio estimator $\hat{\bar{Y}}_{RC}$ and $\hat{\bar{Y}}_{DR_1}$ give slightly different outcomes in biases.

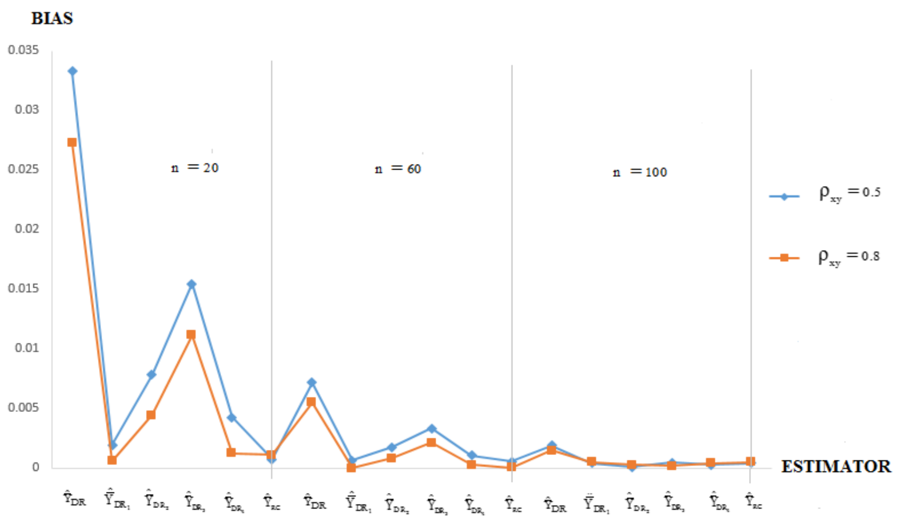

Figure 1. The bias of the proposed estimators and the classical ratio estimator under double sampling for N = 1000, $n' = 200$, $\bar{X} = 2$, $\bar{Y} = 4$ and $V(\bar{X}) = V(\bar{Y}) = 1$.

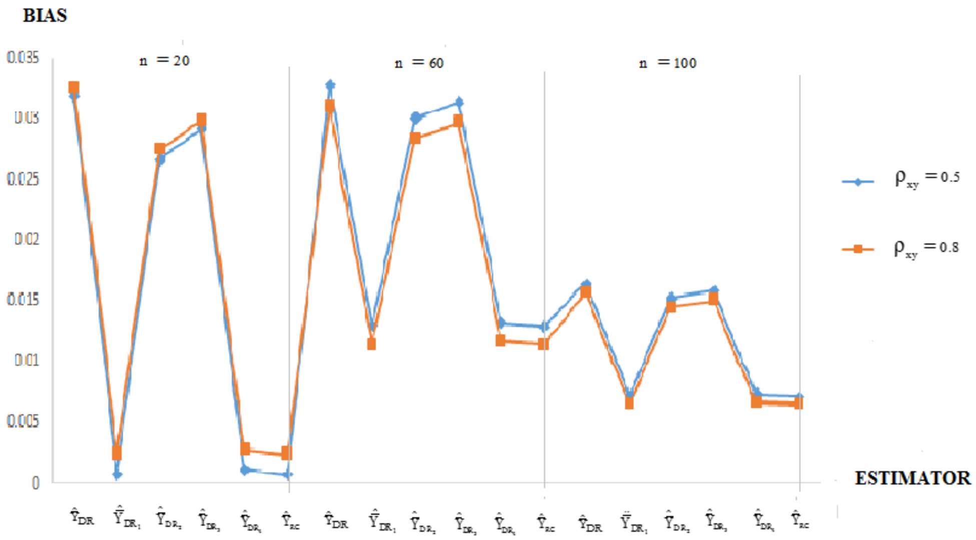

Figure 2. The bias of the proposed estimators and the classical ratio estimator under double sampling for N = 1000, n' = 200, $\bar{X}$ = 200, $\bar{Y}$ = 400 and $V(\bar{X})$ = 100 and $V(\bar{Y})$ = 9.

4. Conclusions

In this paper, we proposed to extend the ratio estimator proposed by Subramani and Kumarapandiyan (2012) using double sampling. Double sampling is used to estimate the unknown population quartile function of an auxiliary variable. We compared the proposed estimators with the classical ratio estimators using double sampling. The simulation results showed that the proposed estimators were more efficient than the classical ratio estimator using double sampling. Therefore, it is a very useful technique that can be used in practice to improve the precision of estimators by using auxiliary variables even where the population values are unknown.

Acknowledgements

This work was funded by The Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand. I would like to thank the unknown referees for offering useful suggestions for improving this paper.

References

Cochran, W. G. (1977), *Sampling techniques* (3rd ed.). New York, NY: John Wiley and Sons.

Neyman, J. (1938). Contribution to the theory of sampling human populations. *Journal of the American Statistical Association, 33*, 101-116.

Pandey, B. N., & Dubey. V. (1988). Modified product estimator using coefficient of variation of auxiliary variate. *Assam Statistical*, 64-66.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika, 43*, 353-360.

Rao, J. N. K., & Sitter, R. R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika, 82*, 453-460.

Singh. H. P., & Tailor. R. (2003). Use of known correlation coefficient in estimating the finite population mean. *Statistics in Transition, 6*, 555-560.

Singh. H. P., & Tailor. R. (2005). Estimation of finite population mean with known coefficient of variation of an auxiliary character. *Statistica, anno LXV, 3*.

Singh, H. P., & Upadhyaya, L. N. (1986). A dual to modified ratio estimator using coefficient of variation of auxiliary variable. *Proceedings of the National Academy of Sciences, India 1986*.

Sisodia, B. V., & Dwivedi, V. K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. *Journal Indian Society of Agricultural Statistics, 33*, 1318.

Subramani, J., & Kumarapandiyan, G. (2012). Modified ratio estimators for population mean using function of quartiles of auxiliary variable. *Bonfring International Journal of Industrial Engineering and Management Science, 2*(2), 19-23.

Yan, Z., & Tian, B. (2010). Ratio method to the mean estimation using coefficient of skewness of auxiliary variable. *ICICA 2010, Part II, CCIS. 106*, 103-110.