

Research Article

Improved ratio estimators of population mean using transformed variable in double sampling

Uraiwan Jaroengertikun¹ and Nuanpan Lawson^{1,*}

¹Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

*E-mail: nuanpan.n@sci.kmutnb.ac.th
Received: 27/06/2018; Accepted: 07/09/2018

Abstract

Transformation techniques can be useful for improving the efficiency of the estimators. New ratio estimators for population mean using both the transformed auxiliary variable and the transformed study variable in double sampling have been proposed. To see how the proposed estimators perform, a simulation study had been conducted in which a percent relative efficiency was used to compare the proposed estimators with some existing estimators. The results showed that the proposed estimators were more efficient than the existing estimators.

Keywords: ratio estimator, auxiliary variable, mean square error, double sampling

Introduction

Using the benefit of known auxiliary variable, related to the study variable can possibly decrease the bias and also increase the efficiency of the estimators. Ratio estimator is one of a well-known estimators that use the information on an auxiliary variable X that related to the study variable Y in order to estimate the population mean of the variable of interest. Cochran (1977) proposed to use an auxiliary variable, related to the variable of interest to improve the population mean estimator in simple random sampling. The classical ratio estimator is given as below.

$$\hat{Y}_R = \frac{\bar{y}}{\bar{x}} \bar{X}, \quad (1)$$

where \bar{y} and \bar{x} are the sample means of the study variable Y and an auxiliary variable X respectively. \bar{X} is the population mean of an auxiliary variable X . The mean square error of the ratio estimator is given as follows.

$$\text{MSE}(\hat{Y}_R) = \left(\frac{1-f}{n} \right) \bar{Y}^2 (C_y^2 + C_x^2 - 2\rho C_x C_y),$$

where C_x and C_y are the coefficient of variation of X and Y respectively. ρ is the correlation between X and Y and a sampling fraction $f = n/N$ where N is the population size and n is the sample size from simple random sampling without replacement.

Many authors have used the ratio estimator to estimate population mean with the use of known population values of auxiliary variables [see, e.g., Sisodia & Dwivedi (1981), Singh & Upadhyaya (1986), Kadilar & Cingi (2004), Nangsue (2009), Yan & Tian (2010), Subramani & Kumarapandiyan (2012) and Soponviwatkul & Lawson (2017)]. Moreover, some authors have applied the transformation method to the auxiliary variable with the purpose of producing a more efficient population mean estimation.

Srivenkataramana (1980) proposed to transform an auxiliary variable x_i to increase the performance of the population mean estimator. The variable x_i is transformed to $x_i^* = (N\bar{X} - nx_i)/(N - n)$, $i = 1, 2, \dots, N$ and a corresponding sample mean of x_i^* is $\bar{x}^* = (N\bar{X} - n\bar{x})/(N - n) = (1 + \pi)\bar{X} - \pi\bar{x}$, where $\pi = n/(N - n)$. By replacing \bar{x}^* into (1), Srivenkataramana (1980) ratio estimator is shown as follows.

$$\hat{Y}_{RS} = \frac{\bar{y}}{\bar{x}^*} \bar{X} \quad (2)$$

The mean square error of this estimator is given as follows.

$$MSE(\hat{Y}_{RS}) = \left(\frac{1-f}{n} \right) \bar{Y}^2 (C_y^2 + (1-\pi^2)C_x^2 - 2(1-\pi)\rho C_x C_y)$$

Later many authors have also improved the efficiency of population mean estimation using the population mean estimator on a study variable by using this variable transformation (see e.g., Tailor & Sharma, 2009; Onyeka et al., 2013). However, most authors only transformed the auxiliary variable X . Later, Adewara et al. (2012) proposed to transform both an auxiliary variable X and a study variable Y to increase the efficiency in estimating the population mean of the variable of interest. They proposed to use a transformed sample of auxiliary variable \bar{x}^* and study variable \bar{y}^* in the classical ratio estimator. The transformed sample mean of the auxiliary and the study variables are $\bar{x}^* = (N\bar{X} - n\bar{x})/(N - n) = (1 + \pi)\bar{X} - \pi\bar{x}$ and $\bar{y}^* = (N\bar{Y} - n\bar{y})/(N - n) = (1 + \pi)\bar{Y} - \pi\bar{y}$ respectively. The ratio estimator after using transformed variable \bar{x}^* and \bar{y}^* of Adewara et al. (2012) is defined as follows.

$$\hat{Y}_{RA} = \frac{\bar{y}^*}{\bar{x}^*} \bar{X}. \quad (3)$$

$$\hat{Y}_{RA_2} = \bar{y}^* \left(\frac{\bar{X} + C_x}{\bar{x}^* + C_x} \right), \quad (4)$$

$$\hat{Y}_{RA_3} = \bar{y}^* \left(\frac{\bar{X} + \rho}{\bar{x}^* + \rho} \right), \quad (5)$$

where C_x and ρ is the population coefficient of variation and correlation coefficient of an auxiliary variable respectively.

However, increasing the efficiency of the ratio estimator by using the transformed study variable which is the function of its population mean is in doubt by some authors (see e.g., Onyeka et al., 2013) because if the population mean of the study variable is known then there is no need to estimate its value. Therefore, it is not useful in practice to apply Adewara et al. (2012) estimator to estimate the population mean of the study variable that is already known.

Neyman (1938) proposed a technique called double sampling which is a very useful sampling technique in estimating these unknown parameters. Let N be the number of units in a population $U = U_1, \dots, U_N$. A simple random sampling without replacement (SRSWOR) is used to collect a large sample of size n' in the first phase sampling to collect the information from an auxiliary variable as it is less time consuming and less cost to obtain when compared to a full study variable. Then a smaller sample size n ($n < n'$) is selected from the second phase using SRSWOR to obtain the information on both the study variable and an auxiliary variable. Assume that \bar{X} in (1) is unknown, the classical ratio estimator under double sampling is given by

$$\hat{Y}_{R.DS} = \frac{\bar{y}}{\bar{x}'} \bar{x}', \quad (6)$$

where \bar{x}' is an unbiased estimator of population mean \bar{X} from the first phase sampling based on sample size n' . \bar{x} and \bar{y} are the sample mean of the auxiliary variable and the variable of interest that have been estimated from the second phase sampling respectively.

The mean square error of this estimator is given as follows.

$$MSE(\hat{Y}_{R.DS}) = MSE(\hat{Y}_R) + \left(\frac{1}{n'} - \frac{1}{n}\right) (2\rho C_x C_y - C_x^2).$$

Later Lawson (2016) proposed new ratio estimators for population mean by adjusting Adewara et al. (2012) estimators in which unknown population means of both an auxiliary variable and the study variable are estimated. The transformed variables of the auxiliary and the study variables are given by

$$x_i^* = \frac{N\bar{x}' - n'x_i}{N - n'} \quad \text{and} \quad y_i^* = \frac{N\bar{y}' - n'y_i}{N - n'}, \quad i = 1, 2, \dots, n'.$$

A sample mean for x_i^* and y_i^* are

$$\bar{x}^* = \frac{N\bar{x}' - n'\bar{x}}{N - n'} = (1 + \pi)\bar{x}' - \pi\bar{x} \quad \text{and} \quad \bar{y}^* = \frac{N\bar{y}' - n'\bar{y}}{N - n'} = (1 + \pi)\bar{y}' - \pi\bar{y}.$$

Assume that the population information of the auxiliary variable and the study variable are unknown. Lawson (2016)'s estimators are defined as below.

$$\hat{Y}_{RT_X} = \frac{\bar{y}}{\bar{x}^*} \bar{x}', \quad (7)$$

$$\hat{Y}_{RTy} = \frac{\bar{y}^*}{\bar{x}} \bar{x}', \quad (8)$$

$$\hat{Y}_{RTxy} = \frac{\bar{y}^*}{\bar{x}^*} \bar{x}', \quad (9)$$

where \bar{x}' is an unbiased estimator of population mean \bar{X} from the first phase sampling based on sample size n' . \bar{x} and \bar{y} are the sample mean of the auxiliary variable and the variable of interest from the second phase sampling respectively. \bar{x}^* and \bar{y}^* are the transformed sample mean of the auxiliary variable and the variable of interest respectively that have been estimated from the second phase of sampling.

In this paper, we aim to propose new ratio estimators for population mean using transformation technique following Lawson (2016) and Adewara et al. (2012) using both the auxiliary variable and the study variable in double sampling. The proposed estimators will be compared with the classical ratio estimator using a percentage of relative efficiency (PRE) in double sampling.

The proposed ratio estimators are presented in Section 2. The efficiency comparison of the proposed estimators and the existing ratio estimators using simulation study are given in Section 3. The conclusions are given in Section 4.

Methods

We propose to adjust Adewara et al. (2012)'s estimator following Lawson (2016)'s estimator by using the transformation technique to transform for both the auxiliary variable and the study variable in double sampling. We have suggested to estimate the population mean and also the population values of the auxiliary variable such as coefficient of variation, correlation coefficient, the quartile functions, median, coefficient of skewness and coefficient of kurtosis based on sample from the first phase sampling.

The proposed estimators $\hat{Y}_{RN_i}, i = 1, \dots, 10$ are given as follows.

$$\hat{Y}_{RN_1} = \bar{y}^* \left(\frac{\bar{x}' + c_x}{\bar{x}^* + c_x} \right), \quad (10)$$

$$\hat{Y}_{RN_2} = \bar{y}^* \left(\frac{\bar{x}' + r}{\bar{x}^* + r} \right), \quad (11)$$

$$\hat{Y}_{RN_3} = \bar{y}^* \left(\frac{\bar{x}' + q_1}{\bar{x}^* + q_1} \right), \quad (12)$$

$$\hat{Y}_{RN_4} = \bar{y}^* \left(\frac{\bar{x}' + q_3}{\bar{x}^* + q_3} \right), \quad (13)$$

$$\hat{Y}_{RN_5} = \bar{y}^* \left(\frac{\bar{x}' + q_r}{\bar{x}^* + q_r} \right), \quad (14)$$

$$\hat{Y}_{RN_6} = \bar{y}^* \left(\frac{\bar{x}' + q_d}{\bar{x}^* + q_d} \right), \quad (15)$$

$$\hat{Y}_{RN_7} = \bar{y}^* \left(\frac{\bar{x}' + q_a}{\bar{x}^* + q_a} \right), \quad (16)$$

$$\hat{Y}_{RN_8} = \bar{y}^* \left(\frac{\bar{x}' + m}{\bar{x}^* + m} \right), \quad (17)$$

$$\hat{Y}_{RN_9} = \bar{y}^* \left(\frac{\bar{x}' + \beta_1}{\bar{x}^* + \beta_1} \right), \quad (18)$$

$$\hat{Y}_{RN_{10}} = \bar{y}^* \left(\frac{\bar{x}' + \beta_2}{\bar{x}^* + \beta_2} \right), \quad (19)$$

where c_x is the sample coefficient of variation, r is the sample correlation coefficient, q_1, q_3, q_r, q_d, q_a are the sample values of the first and the third quartiles, the inter-quartile range, the semi-quartile range and the quartile average, m is median, β_1 and β_2 are coefficient of skewness and coefficient of kurtosis of an auxiliary variable X respectively. These values of an auxiliary variable have been estimated from the first phase sampling.

Therefore, the general form for the proposed estimator \hat{Y}_{RN} is given as follows.

$$\hat{Y}_{RN} = \bar{y}^* \left(\frac{\bar{x}' + a}{\bar{x}^* + a} \right), \quad (20)$$

where a is real numbers or estimated value of parameters of an auxiliary variable from the first phase sampling.

Results and discussion

The population size $N=1,000$ is generated to see the performance of the proposed estimators. A sample of n' ($n' = 50, 300$) units is selected in the first phase sampling then a sample size n ($n = 5, 10, 15$ for $n' = 50$ and $n = 5, 15, 60$ for $n' = 300$) is selected from the second phase sampling with simple random sampling without replacement. The correlation between the variable of interest and an auxiliary variable are $-0.2, -0.5, -0.7, 0.2, 0.5$ and 0.7 . The value of the variable of interest (X, Y) is generated from the bivariate normal distribution with mean equal to 20 and 10 respectively and the variance for both variables is equal to 1 and repeat 10,000 times. A percentage of relative efficiency has been used to compare the

performance of the proposed estimators with the classical ratio estimator under double sampling. The results are presented in Tables 1 to 4.

Table 1. Percent relative efficiency of the proposed estimators over the classical ratio estimator under double sampling for $N = 1000$, $n' = 50$ with a negative relationship between X and Y

Estimator	$\rho = -0.2$			$\rho = -0.5$			$\rho = -0.7$		
	n=5	n=10	n=15	n=5	n=10	n=15	n=5	n=10	n=15
\hat{Y}_{DR}	100	100	100	100	100	100	100	100	100
\hat{Y}_{RTX}	140	137	132	174	174	170	198	201	201
\hat{Y}_{RTY}	427	323	241	569	450	340	685	566	434
\hat{Y}_{RTXY}	1219	522	278	1411	581	299	1533	617	312
\hat{Y}_{RN1}	1219	522	278	1411	582	300	1533	617	312
\hat{Y}_{RN2}	1218	522	278	1407	579	297	1527	612	309
\hat{Y}_{RN3}	1252	549	300	1470	627	335	1612	677	357
\hat{Y}_{RN4}	1252	549	301	1472	629	337	1614	679	358
\hat{Y}_{RN5}	1224	526	281	1419	588	304	1544	625	318
\hat{Y}_{RN6}	1222	524	280	1415	585	302	1539	621	315
\hat{Y}_{RN7}	1252	549	301	1471	628	336	1613	678	358
\hat{Y}_{RN8}	1252	549	301	1471	628	336	1613	678	358
\hat{Y}_{RN9}	1219	522	278	1410	582	300	1533	617	312
\hat{Y}_{RN10}	1228	530	284	1426	593	308	1556	631	322

Tables 1 and 3 show similar results, we can see clearly that the proposed estimators using a transformed variable for both X and Y variables performed a lot better than the classical ratio estimator and Lawson (2016)'s estimators with a negative relationship between X and Y. The proposed estimator using transformed variables for X, Y and $q_3 (\hat{Y}_{RN4})$ performed the best.

All other proposed estimators gave similar PRE except the proposed estimator using a transformed variable for X, Y and $r_X (\hat{Y}_{RN2})$ gave slightly less PRE when compared to \hat{Y}_{RTXY} .

The proposed estimators performed well, especially for a small sampling fraction.

Table 2. Percent relative efficiency of the proposed estimators over the classical ratio estimator under double sampling for $N = 1000$, $n' = 50$ with a positive relationship between X and Y

Estimator	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.7$		
	n=5	n=10	n=15	n=5	n=10	n=15	n=5	n=10	n=15
\hat{Y}_{DR}	100	100	100	100	100	100	100	100	100
\hat{Y}_{RTX}	107	103	55	77	74	71	58	55	97
\hat{Y}_{RTY}	310	223	86	211	152	113	153	180	163
\hat{Y}_{RTXY}	1008	437	189	767	358	213	595	1514	243
\hat{Y}_{RN_1}	1008	438	189	767	358	213	595	1514	243
\hat{Y}_{RN_2}	1008	438	189	767	358	213	595	1513	243
\hat{Y}_{RN_3}	1016	446	180	761	354	210	585	1499	251
\hat{Y}_{RN_4}	1017	446	180	761	357	210	584	1498	251
\hat{Y}_{RN_5}	1011	439	188	767	358	213	594	1513	244
\hat{Y}_{RN_6}	1010	438	189	767	358	213	595	1513	244
\hat{Y}_{RN_7}	1016	446	180	761	354	210	584	1498	251
\hat{Y}_{RN_8}	1016	446	180	761	354	210	584	1498	251
\hat{Y}_{RN_9}	1009	437	190	767	358	213	595	1514	243
$\hat{Y}_{RN_{10}}$	1011	441	188	767	358	213	593	1511	246

Tables 2 and 4 show similar results but the results are different when compared to Tables 1 and 3. We can see that the proposed estimators using a transformed variable for X and Y variables and some sample of an auxiliary variable such as q_1, q_3, q_a and median performed better than the classical ratio and Lawson (2016)'s estimators when $\rho = 0.2$. When $\rho = 0.5$ and 0.8 , all new proposed estimators performed better than the classical ratio estimator and some of them gave the same PRE when compared to Lawson (2016)'s estimator using a transformed variable for both X and Y variables (\hat{Y}_{RTXY}). The proposed estimators performed well, especially for a small sampling fraction and when there was a weak positive correlation between X and Y.

Table 3. Percent relative efficiency of the proposed estimators over the classical ratio estimator under double sampling for $N = 1000$, $n' = 300$ with a negative relationship between X and Y

Estimator	$\rho = -0.2$			$\rho = -0.5$			$\rho = -0.7$		
	n=5	n=15	n=60	n=5	n=15	n=60	n=5	n=15	n=60
\hat{Y}_{DR}	100	100	100	100	100	100	100	100	100
\hat{Y}_{RTX}	144	141	143	176	173	181	197	196	210
\hat{Y}_{RTY}	529	541	369	661	691	530	756	800	682
\hat{Y}_{RTXY}	11463	3646	623	13820	4297	692	15404	4715	733
\hat{Y}_{RN1}	11463	3646	623	13820	4298	692	15406	4716	733
\hat{Y}_{RN2}	11461	3644	622	13811	4290	688	15389	4701	726
\hat{Y}_{RN3}	11537	3719	669	13960	4426	765	15599	4886	826
\hat{Y}_{RN4}	11539	3721	669	13964	4430	768	15605	4892	829
\hat{Y}_{RN5}	11474	3657	630	13840	4315	702	15432	4739	745
\hat{Y}_{RN6}	11469	3651	627	13830	4307	697	15419	4727	739
\hat{Y}_{RN7}	11538	3720	669	13962	4428	767	15602	4889	828
\hat{Y}_{RN8}	11538	3720	669	13962	4428	767	15602	4889	828
\hat{Y}_{RN9}	11463	3646	624	13820	4297	692	15405	4715	732
\hat{Y}_{RN10}	11484	3666	636	13857	4330	710	15455	4758	755

We can see clearly that the proposed estimators performed very well when the sample size n is small in Tables 1 to 4 when there is both a negative relationship and positive relationship between X and Y. As you can see the PREs decreased when the sample size n increased so the new transformation method can be used to increase the efficiency of the population mean estimator with a small sampling fraction which is a benefit to researchers in saving money. However, the new estimators seem to be less efficient when the sample size is larger.

Table 4. Percent relative efficiency of the proposed estimators over the classical ratio estimator under double sampling for $N = 1000$, $n' = 300$ with a positive relationship between X and Y

Estimator	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.7$		
	n=5	n=15	n=60	n=5	n=15	n=60	n=5	n=15	n=60
\hat{Y}_{DR}	100	100	100	100	100	100	100	100	100
\hat{Y}_{RTX}	106	106	99	77	75	71	57	55	54
\hat{Y}_{RTY}	430	363	236	303	250	157	220	180	113
\hat{Y}_{RTXY}	9169	2708	526	6655	2003	433	4928	1514	361
\hat{Y}_{RN_1}	9169	2708	526	6655	2003	433	4928	1514	361
\hat{Y}_{RN_2}	9169	2709	527	6654	2003	433	4927	1513	360
\hat{Y}_{RN_3}	9190	2729	539	6644	1997	426	4908	1499	344
\hat{Y}_{RN_4}	9190	2729	539	6643	1997	426	4907	1498	344
\hat{Y}_{RN_5}	9173	2712	529	6654	2003	433	4926	1513	359
\hat{Y}_{RN_6}	9171	2710	527	6654	2003	433	4927	1513	360
\hat{Y}_{RN_7}	9190	2729	539	6643	1997	426	4907	1498	344
\hat{Y}_{RN_8}	9190	2729	539	6643	1997	426	4907	1498	344
\hat{Y}_{RN_9}	9170	2708	526	6655	2003	433	4928	1514	361
$\hat{Y}_{RN_{10}}$	9177	2715	530	6654	2003	433	4924	1511	358

Conclusion

Transformation method has been used to improve the efficiency of the population mean estimator using ratio estimation under double sampling. In this paper, we proposed to adjust the transformed variable technique proposed by Adewara et al. (2012) following Lawson (2016)'s estimator by using the transformation technique to transform for both the auxiliary variable and the study variable in double sampling. The proposed estimators using a transformed variable performed better than the classical ratio estimator and Lawson (2016)'s estimator under double sampling. We can see a huge improvement in the percentage of relative efficiency when using the transformed variables for a small sampling fraction. All proposed estimators performed well when the relationship between X and Y variable was negative and when there was a weak positive correlation between X and Y . Therefore, it is a very useful technique that can be used in practice to improve the precision of estimators by using both auxiliary and study variables and also some estimated values of the auxiliary variable when the population values are not known in a small sample.

Acknowledgements

This work was funded by The Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand. Contract no. 6145103. We would like to thank the unknown referees for offering useful suggestions for improving this paper.

References

- Adewara, A. A., Singh, R. & Kumar, M. (2012). Efficiency of some modified ratio and product estimators using known value of some population parameters. *International Journal of Applied Science and Technology*, 2(2), 76-79.
- Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). New York: John Wiley and Sons.
- Kadilar, C. & Cingi, H. (2004). Ratio Estimators in simple random sampling. *Applied Mathematics and Computation*, 151(3), 893-902. [https://doi.org/10.1016/S0096-3003\(03\)00803-8](https://doi.org/10.1016/S0096-3003(03)00803-8)
- Lawson, N. (2016). Estimating population mean using transformed variables in double sampling. *Full Paper Proceeding ECBA-2016*, 158(8), 1-5. From Hongkong 8th International Conference on "Engineering & Tecnnology, Computer, Basic and Applied Science (ECBA) 2016, 5-6 May 2016, Hongkong.
- Nangsue, N. (2009). Adjusted ratio and regression type estimators for estimation of population mean when some observations are missing. *International Scholarly and Scientific Research & Innovation*, 3(5), 334-337.
- Neyman, J. (1938), Contribution to the theory of sampling human populations, *Journal of the American Statistical Association*. 33, 101-116. <https://doi.org/10.2307/2279117>
- Onyeka, A. C., Nlebedim, V. U. & Izunobi, C. H. (2013). Estimation of population ratio in simple random sampling using variable transformation. *Global Journal of Science Frontier Research Mathematics and Decision Sciences*, 13(4), 57-65.
- Singh, H. P. & Upadhyaya, L. N. (1986). A dual to modified ratio estimator using coefficient of variation of auxiliary variable. *Proceedings National Academy of Sciences India*, 56(A), Part 4, 336-340.
- 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(2), 13-18.
- Soponviwatkul, K. & Lawson, N. (2017). New ratio estimators for estimating population mean in simple random sampling using a coefficient of variation, correlation coefficient and a regression coefficient. *Gazi University Journal of Science*, 30(4), 610-621.
- Srivenkataramana, T. (1980). A dual to ratio estimator in sample surveys. *Biometrika*, 67(1), 199-204. <https://doi.org/10.1093/biomet/67.1.199>
- Subramani, J. & Kumarapandiyam, 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. doi: 10.9756/BIJIEMS.1356
- Taylor, R. & 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*, 10(1), 15-24.
- Yan Z. & Tian B. (2010). Ratio method to the mean estimation using coefficient of skewness of auxiliary variable. In R. Zhu, Y. Zhang, B. Liu & C. Liu (Eds.), *Information Computing and Applications. ICICA 2010. Communications in Computer and Information Science*, vol 106, pp. 103-110. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-16339-5_14