Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/030109

Two Auxiliary Variables in Ratio Method of Estimation and its Application

Mariam Al-Mannai *

Department of Mathematics, College of Science, University of Bahrain, Kingdom of Bahrain
**Email: malmannai@uob.edu.bh*

Received: 17 May. 2013, Revised: 21 Sep. 2013, Accepted: 25 Oct. 2013
Published online: 1 Mar. 2014

Abstract: In this paper an attempt is made to provide a ratio estimator based on two auxiliary variables to estimate the mean of a survey variable when the means as well as the coefficients of variation of two auxiliary variables are known. It is an extension of [1] ratio estimator based on two auxiliary variables, in the sense that the assumptionof the coefficients of correlation between variable under study (y) with auxiliary variables equal [$\rho_{01} = \rho_{02}$] is relaxed. The mean square error (variance) of the suggested estimator, up to terms of order n^{-1}, is obtained. The usefulness of the estimator is demonstrated with the help of examples taken from the literature. The performance of the suggested estimator with that of Olkin's estimator is demonstrated empirically. Since the suggested estimator involves unknown population parameter d (= $\frac{\rho_{01}}{\rho_{02}}$), while in Olkin's estimator w_1 (optimum)requires the knowledge of unknown population parameters such as ρ_{01}, C_y and ρ_{02}, hence the performance is also studied by deviating d and w_1 from 0% to 30 % in either directions.

Keywords: Auxiliary Variable; Mean Square Error; Multivariate Ratio Estimator.

1. Introduction

It is known that the proper use of auxiliary variable(s), in simple random sampling with or without replacement, may lead to more efficient estimators under certain conditions (ratio, product and regression methods are widely used in real life surveys).Olkin [2] introducedthe use of several auxiliary variables, each positively correlated with the study variable, in the ratio method of estimation by considering a linear combination of ratio estimators based on each auxiliary variable separately. Raj [3] suggested a method for using multi-auxiliary variables through a linear combination of single difference. Since then a number of researchers have contributed on multivariate generalizations of ratio estimators or multivariate extension on difference estimators.Interested readers may refer to ([4], [5], [6], [7], [1], [8], [9], [10], [11], [12],[13], [14], [15], [16], [17], [18], [19], [20],[21], [22]). It is worth to mention that, in spite of the wide variety of proposals, most of the above estimators exhibit the same performance at the first order of approximation.

Many survey practitioners believe that Olkin's multivariate ratio estimator couldn't be used in real life surveys because it requires the knowledge of many unknown population parameters. Cochran (1977 p 185, [23]) has mentioned that the unknown population parameters required in weights ofratio estimator could be used from sample values. An empirical study based on 16 populations is made by [1], and found Olkin's multivariate ratio estimator to be robust against deviations from the optimal conditions. Therefore, one can use Olkin's multivariate ratio estimator in actual surveys by using guess values based on earlier surveys or from the knowledge of the experts in the field. The estimator suggested by [1] assumes the coefficients of correlation between variable under study (y) with auxiliary variables are equal, which restricts its use in real life situations for which the condition doesn't hold. Further, they showed that the expressions of bias and variance, to the first degree of approximation, are identical to that of Olkin's ratio estimator based on two auxiliary variables under the conditions that coefficients of variation of both the auxiliary variables (say x_1 and x_2), and also the coefficients of correlation between variable under study (y) with auxiliary variables are equal. In this paper we extend the estimator suggested by [1] without the assumptions thatneither the coefficients of variation of the auxiliary variables to be equal nor the coefficients of correlation between variable under study (y) with auxiliary variables should be same. The expressions of bias and variance of the estimator, to the first degree of approximation, are given for the estimator. The performance of the suggested estimator over Olkin's estimator is studied for a wide variety of populations related to medical science, taken from the literature [24] and [25].

2. Notations and Preliminaries:

Consider a finite population U of size N identifiable, distinct units $u_1, u_2, u_i, \ldots\ldots u_N$, It is assumed that study variables y, and auxiliary variables x_j, j= 1, 2 are defined on U.

A simple random sample without replacement of size n is selected from finite population of size N. Let,

$$\bar{y} = n^{-1}\sum_{i=1}^{n} y_i \;;\; \bar{Y}_j = N^{-1}\sum_{i=1}^{N} y_i \;;\; V(y) = \sigma_y^2 = N^{-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2 \;;\; \text{j} = 1, 2.$$

$$\bar{x}_j = n^{-1}\sum_{i=1}^{n} x_i \;;\; \bar{X}_j = N^{-1}\sum_{i=1}^{N} x_i \;;\; V(x_j) = \sigma_{x_j}^2 = N^{-1}\sum_{i=1}^{N}(x_{ji} - \bar{X}_j)^2 \;\text{j=1, 2.}$$

$$C_y = \frac{\sigma_y}{\bar{Y}} \;;\; C_{x_j} = \frac{\sigma_{x_j}}{\bar{X}_j} \;;\; \text{d} = \frac{\rho_{01}}{\rho_{02}}$$

$$\bar{y}_{R_j} = \frac{\bar{y}}{\bar{x}_j}\bar{X}_j \;;\; Cov(y, x_j) = \sum_{i=1}^{N}(y_i - \bar{Y})(x_{ji} - \bar{X}_j) \;;\quad \text{j} = 1, 2$$

$$\rho_{0j} = \frac{Cov(y, x_j)}{\sqrt{V(y)V(x_j)}} \;;\quad \text{j} = 1, 2.; \rho_{12} = \frac{Cov(x_k, x_l)}{\sqrt{V(x_k)V(x_l)}} ; k \neq l$$

$$V(y_{R_j}) = \frac{N-n}{Nn}\bar{Y}^2[C_y^2 + C_{x_j}^2 - 2\rho_{0j}C_yC_{x_j}] \quad ; \text{j} = 1,2$$

$$Cov(\bar{y}_{R_1}, \bar{y}_{R_2}) = \frac{N-n}{Nn}\bar{Y}^2[C_y^2 + \rho_{12}C_{x_1}C_{x_2} - \rho_{01}C_yC_{x_1} - \rho_{02}C_yC_{x_2}]$$

3. Ratio Estimator based on two auxiliary variables:

For the sake of brevity and ready reference, we define Olkin's multivariate ratio estimator $\bar{y}_w$ and the estimator $\bar{y}_{AK}$ suggested by [1] as follows:

$$\bar{y}_w = w_1\bar{y}_{R_1} + w_2\bar{y}_{R_2}; \quad w_1 + w_2 = 1 \tag{1}$$

$$V(\bar{y}_w) = w_1^2V(\bar{y}_{R_1}) + w_2^2V(\bar{y}_{R_1}) + 2w_1w_2Cov(\bar{y}_{R_1}, \bar{y}_{R_1}) \tag{2}$$

The optimum weights which maximizes the precision of $\bar{y}_w$, are

$$w_1 = \frac{C_{x_2}^2 - \rho_{02}C_yC_{x_2} + \rho_{01}C_yC_{x_1} + \rho_{12}C_{x_1}C_{x_2}}{C_{x_1}^2 + C_{x_2}^2 - 2\rho_{12}C_{x_1}C_{x_2}} \tag{3}$$

w_1 requires the knowledge of unknown population parameters such as ρ_{01}, C_y and ρ_{02}.

Using two auxiliary variables, [1] considered different approach in ratio method of estimation and defined the estimator for those populations for which $\rho_{01} = \rho_{02} = \rho$ as:

$$\bar{y}_{AK} = \bar{y}[(\frac{\sigma_{x_1}}{\sigma_{x_2}}\bar{X}_1 + \bar{X}_2)/(\frac{\sigma_{x_1}}{\sigma_{x_2}}\bar{x}_1 + \bar{x}_2)] \tag{4}$$

The expressions of bias and variance, to the first degree of approximations are:

$$\text{Bias}(\bar{y}_{AK}) = \frac{\bar{Y}}{n}(\frac{N-n}{N})[2C_{x_1}C_{x_2}/(C_{x_1} + C_{x_2})\{C_{x_1}C_{x_2}(1+\rho_{12})/(C_{x_1} + C_{x_2}) - \rho C_y\}] \tag{5}$$

$$\text{V}(\bar{y}_{AK}) = \frac{\bar{Y}^2}{n}(\frac{N-n}{N})[C_y^2 + 2C_{x_1}C_{x_2}/(C_{x_1} + C_{x_2})\{C_{x_1}C_{x_2}(1+\rho_{12})/(C_{x_1} + C_{x_2}) - 2\rho C_y\}] \tag{6}$$

Under the assumptions $\rho_{01} = \rho_{02} = \rho$ and $C_{x_1} = C_{x_2} = C_x$, the expressions of bias and variance reduce to

$$\text{Bias}(\bar{y}_{AK}) = \frac{(N-n)\bar{Y}}{Nn}(C_x^2 - \rho C_xC_y) \tag{7}$$

$$\text{V}(\bar{y}_{AK}) = \frac{(N-n)\bar{Y}^2}{Nn}[C_y^2 - 2\rho C_xC_y + \frac{C_x^2}{2}(1+\rho_{12})] \tag{8}$$

The expressions are same as that of Olkin'smultivariate ratio estimator based on two auxiliary variablesunder the assumptions $\rho_{01} = \rho_{02}$ and $C_{x_1} = C_{x_2}$.

It is worth to mention that most of the ratio estimators based on two auxiliary variables require the knowledge of unknown population parameters, andexhibit the same performance at the first order of approximation.In order to use them in actual surveys, Cochran (1977 p 185) has mentioned that the unknown population parameters required in weights of ratio estimator could be used from sample values. Alternatively, the information is available, based on earlier surveys or from the knowledge of the experts in the field could be used.

4. Modified Agarwal and Kashyap's Ratio Estimator based on two auxiliary variables:

In this section we redefine [1] ratio estimator based on two auxiliary variables without the assumptions $\rho_{01} = \rho_{02}$, so that it has wider applicability in real situations. The estimator is:

$$\bar{y}_{AKM} = \bar{y}[(\frac{k_1}{k_2}\bar{X}_1 + \bar{X}_2)/(\frac{k_1}{k_2}\bar{x}_1 + \bar{x}_2] \qquad (9)$$

where $\frac{k_1}{k_2} = (\frac{\sigma_{x_1}}{\sigma_{x_2}})(\frac{d-\rho_{12}}{1-d\,\rho_{12}})$;

The expressions of bias and variance of equ (9) are:

$$\text{Bias}\,(\bar{y}_{AKM}) = \frac{(N-n)\bar{Y}}{nN}[(\frac{d^2-2d\rho_{12}+1}{\frac{d-\rho_{12}}{C_{x_1}}+\frac{1-d\rho_{12}}{C_{x_2}}})(\frac{1-\rho_{12}^2}{\frac{d-\rho_{12}}{C_{x_1}}+\frac{1-d\rho_{12}}{C_{x_2}}} - \rho_{02}C_y)] \quad (10)$$

$$\text{V}(\bar{y}_{AKM}) = \frac{(N-n)\bar{Y}^2}{nN}[C_y^2 + (\frac{d^2-2d\rho_{12}+1}{\frac{d-\rho_{12}}{C_{x_1}}+\frac{1-d\rho_{12}}{C_{x_2}}})(\frac{1-\rho_{12}^2}{\frac{d-\rho_{12}}{C_{x_1}}+\frac{1-d\rho_{12}}{C_{x_2}}} - 2\rho_{02}C_y)] \quad (11)$$

5. Empirical Results

In this section, we study the performance of $\bar{y}_{AKM}$ with that ofOlkin'sestimator $\bar{y}_w$. We used only Olkin's estimator for comparison, because in the past many authors have made empirical studies by comparing $\bar{y}_w$ with $\bar{y}_{R_j}$; j =1, 2. The performance of $\bar{y}_{AKM}$ over $\bar{y}_w$ is also studied by deviating actual value of d as well as that of w_1 from optimum values between0% to 30%in the interval of 10% on either side for several known populations taken from the literature [24, 25] whose characteristics are given in Table-1. It is worth to mention that [1], made an empirical study for 16 populations, all having ρ_{01} and ρ_{02} close to each other, and found Olkin's multivariate ratio estimator to be robust against deviations from the optimal conditions. Since the modified estimator in section 4 is suggested for those populations for which most of the populations do not have ρ_{01} and ρ_{02} close to each other, hence empirical robustness of $\bar{y}_w$ is also studied by deviating w_1 from optimum values on either side.

Table1: Characteristics of the populations

Number	N	C_y	C_{x_1}	C_{x_2}	ρ_{01}	ρ_{02}	ρ_{12}	$w_1(opt)$	(actual) d
1	35	13.85	6.03	16.29	0.444	0.878	0.067	0.339	0.505
2	100	20.06	20.16	15.30	0.380	0.502	0.456	0.259	0.757
3	30	18.24	17.98	14.93	0.879	0.587	0.666	0.914	1.496
4	20	40.48	41.48	48.37	0.615	0.731	0.148	0.475	0.842
5	15	17.30	15.53	38.18	0.943	0.576	0.495	0.933	1.638
6	25	28.29	40.88	56.11	0.418	0.763	0.450	0.504	0.548
7	47	23.34	35.92	56.53	0.405	0.455	0.507	0.785	0.890
8	51	32.18	31.54	57.37	0.411	0.520	0.457	0.730	0.790
9	40	23.01	35.75	57.17	0.391	0.401	0.494	0.812	0.976
10	50	32.16	31.77	55.07	0.352	0.444	0.381	0.716	0.794
11	70	28.02	36.94	59.87	0.442	0.507	0.482	0.755	0.872
12	90	28.40	36.19	58.68	0.360	0.499	0.443	0.711	0.722
min	15	13.85	6.03	14.93	0.352	0.401	0.067	0.259	0.505
max	100	40.48	41.48	59.87	0.943	0.878	0.666	0.933	1.638
Q_1	29	19.60	19.61	32.71	0.388	0.488	0.427	0.497	0.748
median	43	25.68	33.76	55.59	0.415	0.514	0.456	0.723	0.818
Q_3	56	29.34	36.38	57.22	0.487	0.623	0.494	0.792	0.912

Table-1gives the characteristics of the populations such as population size N, coefficients of variation of the study variable (y), the auxiliary variables x_1 and x_2, the correlation coefficients between (y, x_1) and (y, x_2) actual values of d, and w_1. Since one of the aim of the paper is to study the empirical robustness of Olkin's estimator, hence only those populations are considered which not only have wide range of w_1, but d is away from unity.In table -1, the population size variesfrom 15 to 100, the coefficient of variation of y from 13.85 % to 40.48%, the coefficient of variation of x_1 from 6.03% to 41.48%, the coefficient of variation of x_2 from 14.93% to 59.87%. The correlation coefficient between (y, x_1) varies from 0.35 to 0.94, while the correlation coefficient between (y, x_2) varies from0.40 to 0.88. The optimum values of w_1 varies from 0.259 to 0.933, while d varies from 0.505 to 1.638. It can be seen that except for population number 9, for rest of the populations d is far away from unity. The above described populations thus represent a wide variety of situations.

Table 2: Percentage relative efficiency of suggested estimator $\bar{y}_{AKM}$ over Olkin's estimator $\bar{y}_w$ when optimum values w_1 and true value " d " deviated from 0% to 30%

number							
	0%	**-10%**	**10%**	**-20%**	**20%**	**-30%**	**30%**
1	110.36	143.73	86.94	195.71	69.48	288.43	55.94
2	444.49	361.83	537.39	288.50	642.63	224.35	764.00
3	99.42	90.37	124.07	87.92	189.67	87.07	343.85
4	846.86	324.93	440.60	284.90	259.81	109.41	103.20
5	1118.59	3965.74	110.26	2290.69	494.38	1484.59	741.98
6	112.61	87.64	141.12	66.37	172.88	48.80	207.57
7	133.36	122.92	135.09	104.93	128.52	82.13	115.13
8	164.24	138.93	181.05	109.02	186.38	78.72	178.50
9	146.34	143.35	139.02	129.57	123.44	106.79	102.38
10	209.92	188.09	221.81	158.86	222.41	125.69	211.48
11	158.21	142.06	165.11	118.63	162.29	91.28	150.48
12	254.91	199.04	313.36	148.21	372.06	104.50	429.06
min	99.42	87.64	86.94	66.37	69.48	48.80	55.94
max	1118.59	3965.74	537.39	2290.69	642.63	1484.59	764.00
Q_1	128.17	134.93	132.34	108.00	153.84	85.83	112.15
median	161.23	143.54	153.11	138.89	188.02	105.64	193.04
Q_3	302.30	230.52	244.70	218.01	287.87	150.36	365.15

Table 2 gives the percentage relative efficiency of modified ratio estimator over Olkin's ratio estimator $[=\frac{V(\bar{y}_w)}{V(\bar{y}_{AKM})}]*100$, using two auxiliary variables, and also deviating optimum values w_1 and d by 0% to 30% from the true values. It can be seen that by using true value of d, and optimum value w_1, the performance of $\bar{y}_{AKM}$ is much better than that of $\bar{y}_w$.When the actual values of w_1 and d,are deviated on upper side from 10% to 30 %, the gain in efficiency of $\bar{y}_{AKM}$ over $\bar{y}_w$ is considerable. When w_1 and d, are deviated on lower side from -10% to -20%, the performance of $\bar{y}_{AKM}$ over $\bar{y}_w$ is better for most of the populations. This study of empirical robustness is important because the true values of w_1 and d are rarely available in actual surveys, and if survey practitioners follow Cochran (1977 p 185) about the unknown population parameter w_1 and d,one should know how much gain or loss is expected. It is worth to mention that [1] in their empirical study found $\bar{y}_w$ to be robust, for those populations for which d are close to unity. That is the reason that in the present empirical study only those populations (except number 9) are considered for which d are not close to unity, but varies considerably, and away from unity. For this set of populations $\bar{y}_w$ is not robust and hence contradicts the observations made by [1].

Acknowledgments:

The author is thankful to the referee and the Associate Editor for the comments which helped to improve the paper.

References:

[1] Agarwal, S.K. and Kashyap, P. Multivariate ratio estimator: Theory and Practice. Cahiers du CERO (Statistics & Operations Research), **35, 1-2,** 3-13 (1993).

[2] Olkin, I. Multivariate ratio estimation for finite populations, Biometrika, **45,** 154–165 (1958).

[3] Raj, D. On a method of using multi-auxiliary information in sample surveys, Journal of the American Statistical Association, **60**, 154–165 (1965).

[4] Rao, P. S. R. S. and Mudholkar, G. S. Generalized multivariate estimator for the mean of finite populations, Journal of the American Statistical Association, **62,** 1009–1012 (1967).

[5] John, S. On multivariate ratio and product estimators, Biometrika, **56**, 533–536 (1969).

[6]Agarwal, S.K. Two auxiliary variates in ratio method of estimation.Biometrical J., **22-7**, 569-573 (1980).

[7] Agarwal, S.K. and Kumar, P. On two auxiliary variates in ratio method of Estimation. Biometrical J., **27,** 237-238 (1985)

[8] Tracy, D. S., Singh, H. P., and Singh, R. An alternative to the ratio-cum-product estimator in sample surveys.Journal of Statistical Planning Inference, **53,** 375–387 (1996).

[9] Ceccon, C. and Diana, G. Classi di stimatori di tiporapporto in presenza di pi´uvariabiliausiliarie, Proceedings of Italian Statistical Society, XXXVIII Scientific Meeting, Rimini, 9-13 April 1996, 119–126, (1996).

[10] Agarwal, S. K., Sharma, U. K., and Kashyap, S. A new approach to use multivariateauxiliary information in sample surveys, Journal of Statistical Planning and Inference, **60,** 261–267 (1997).

[11] Kadilar, C. and Cingi, H. Estimator of a population mean using two auxiliary variables in simple random sampling, International Mathematical Journal, **5,** 357–367 (2004).

[12] Kadilar, C. and Cingi, H. A new estimator using two auxiliary variables, Applied Mathematics and Computation, **162**, 901–908 (2005).

[13] Kadilar and Cingi, H. Improvement in estimating the population mean in simple random sampling, Applied Mathematics Letter**, 19**, 75–79 (2006).

[14] Perri, P. F. Alcuneconsiderazionisull'efficienzadeglistimatorirapporto-cum-prodotto, Statistica & Applicazioni, **2,** 59–75 (2004).

[15] Perri, P. F. Combining two auxiliary variables in ratio-cum-product type estimators, Proceedings of Italian Statistical Society, Intermediate Meeting on Statistics and Environment, Messina, **21-23**, 193–196 (2005).

[16] Diana, G. and Perri, P. F. Multivariate regression-type estimators, Proceedings of Italian Statistical Society, XLIII Scientific Meeting, Torino, **14-16**, 61–64 (2006).

[17] Diana, G. and Perri, P. F. Estimation of finite population mean using multi-auxiliary information; METRON - International Journal of Statistics, **LXV,** 99-112 (2007).

[18] Gupta, S and Shabbir, J. On improvement in estimating the population mean in simple random sampling, Journal of Applied Statistics, **35**, (2008).

[19] Koyuncu, N. and Kadilar, C. Family of estimators of population mean using two auxiliary variables in stratified random sampling,Communications in Statistics Theory and Methods, **38,** 2398–2417 (2009).

[20] Singh, R, Chauhan, P., Sawan, N. and Smarandache, F.Improvement in estimating population mean using two auxiliary variables, Italian Journal of Pure and Applied Mathematics, **n. 28,** 135-142 (2011).

[21] Chris J. Book review: 'advances in sampling theory - ratio method of estimation' by Cingi, H. and Kadilar, *C.* Journal of the American Statistical Association, **106**, 375-382. ISSN 0162-1459, (2011).

[22] Malik, M. and Singh, R. Some Improved Multivariate-Ratio-Type Estimators Using Geometric and Harmonic Means in Stratified Random Sampling. ISRN Probability and Statistics, Article ID509186, (2012).

[23] Cochran, W. G. *Sampling Techniques*, JohnWiley& Sons, New York, NY, USA, (1977).

[24] Ott, L. *An Introduction to Statistical Methods and Data Analysis*, 2nd Ed**.**Duxbury, Boston. (1984).

[25] Daniel, W. W. *Biostatistics: A Foundation for Analysis in the Health Sciences*, 8th Edition, Wiley, USA, (2005).