IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

Log type direct estimators for domain mean using bivariate auxiliary information

Shashi Bhushan[1], Anoop Kumar *[2] and Rohini Pokhrel[3]

[1]Department of Statistics, University of Lucknow, Lucknow, India, 226007

[2]Department of Statistics, Amity University, Lucknow, U.P., India, 226028

[3]Department of Mathematics & Statistics, Dr. Shakuntala Misra National Rehabilitation University, Lucknow, India, 226017

Abstract: This article suggests a log type direct estimator for domain mean using two auxiliary variables under simple random sampling (SRS). The mean square error expression of the suggested estimator is obtained to the first order of approximation. The performance of the proposed estimator is compared with the existing conventional estimators. An empirical study using real data is conducted to show that the proposed estimator is more efficient than the existing conventional estimators.

Keywords: Small area estimation, direct estimator, bivariate auxiliary information, simple random sampling.

1. Introduction

In order to estimate the parameters of the population under consideration, sample surveys are often undertaken, but when we are interested in the estimation of parameters of a sub-population (domain), it is necessary to have a sufficient number of sampling units in the population. For instance, estimates for the national and state levels are provided through numerous socioeconomic surveys, crop estimation, and health surveys done by different organizations. Nevertheless, due to the insufficient sample size, estimate of parameters is not offered for district and sub-district levels. Small sample sizes for estimate will result in significant standard errors, which might result in inaccurate parameter estimation. The small area estimate (SAE) approach is employed in this situation for estimating purposes. Due to demands for trustworthy small area estimates from both the public and business sectors, SAE plays a very significant role in sample surveys. The goal of sample surveys, whether they are carried out by government agencies or by for-profit businesses, is to get relatively reliable direct estimates for the character being studied at both the population and subpopulation levels. If the data from the realm of interest are not sufficient to provide “direct estimates” with sufficient precision, the region is deemed “small”. This is as a result of domain-specific direct estimators' excessively high coefficient of variation being provided by fewer sample sizes. Small domain is another term for a small area with small samples.

The suggested estimator for domain mean uses an auxiliary character with an unknown auxiliary mean due to the growing use of domain mean estimation in both the public and commercial sectors. Sample surveys are frequently used in practice to offer estimates for the target population as a whole as well as for specific subpopulations (domains) including age groups, sexes, educational levels, industry sectors, and racial groupings, among others. Instead of using demographic characteristics as small areas, such as village or a census division, it is preferable to use geographic areas.

In his book, Rao (2003) provides illustrations of the estimator based solely on the direct technique. Using auxiliary information, several authors proposed modified and improved estimators for domain mean, namely, Gonzalez (1973), Tikkiwal and Ghiya (2000), Rai and Pandey (2013) and Khare and Ashutosh (2018), among many others in sample surveys. In this article, we propose log type direct estimators for domain mean using bivariate auxiliary information.

Let a finite population consists of 'A' non-overlapped small areas, i.e., domains A_a of size N_a for which estimates are needed. The domains might be various and could constitute small areas of a sampled population, such as a district, tehsil, or other state level subdivision, depending on the situation. Let y denotes the characteristic under study. Further, assume that the auxiliary information is also available and denoted by x, and z. A simple random sample $t_{m,a}^{d} = \bar{y}_a$ $s = (s_1, s_2, ..., s_n)$ of size n is chosen without replacement such that n_a, a=1,2,...,A units in the sample s come from the small area A. As a result, $\sum_{a=1}^{A} N_a = N$ and $\sum_{a=1}^{A} n_a = n$. Let $\bar{X}$ be the population mean of auxiliary characteristic x based on N observations; $\bar{X}_a$ be the population mean of characteristic x for a small domain based on N_a observations; $\bar{Z}$ be the population mean of auxiliary characteristic z based on N observations; $\bar{Z}_a$ be the population mean of characteristic z for a small domain based on N observations; $\bar{x}$ be the sample mean based on n observations on characteristic x; $\bar{x}_a$ be the sample mean based on n_a observations on x; $\bar{z}$ be the sample mean based on n observations on characteristic z; $\bar{z}_a$ be the sample mean based on n_a observations on z; $\bar{Y}$ be the population mean based on N observations on y; $\bar{Y}_a$ be the population mean of domain a based on N_a observations on y; $\bar{y}$ be the sample mean based on n observations on y; $\bar{y}_a$ be the sample mean of domain a based on n_a observations on y.

To establish the properties of the direct estimators, we assume the following notations:
$\bar{y}_a = \bar{Y}_a(1+e_0)$, $\bar{x}_a = \bar{X}_a(1+e_1)$, $\bar{z}_a = \bar{Z}_a(1+e_2)$ provided that $E(e_i) = 0$, $|e_i| < 1$; $i = 0,1,2$, $E(e_0^2) = f_a C_{y_a}^2$, $E(e_1^2) = f_a C_{x_a}^2$, $E(e_2^2) = f_a C_{z_a}^2$, and $E(e_0 e_1) = f_a \rho_{y_a x_a} C_{y_a} C_{x_a}$, $E(e_0 e_2) = f_a \rho_{y_a z_a} C_{y_a} C_{z_a}$
$E(e_1 e_2) = f_a \rho_{x_a z_a} C_{x_a} C_{z_a}$, where $f_a = (N_a - n_a)/N_a n_a$, $S_{x_a}^2 = (N_a - n_a)^{-1} \sum_{i=1}^{N_a} (X_{ai} - \bar{X})^2$,
$S_{y_a}^2 = (N_a - n_a)^{-1} \sum_{i=1}^{N_a} (Y_{ai} - \bar{Y})^2$, $C_{x_a} = S_{x_a}/\bar{X}_a$, $C_{y_a} = S_{y_a}/\bar{Y}_a$, $S_{x_a y_a} = (N_a - 1)^{-1} \sum_{i=1}^{N_a} (X_{ai} - \bar{X})(Y_{ai} - \bar{Y})$, and
$C_{y_a x_a} = \rho_{y_a x_a} C_{y_a} C_{x_a}$.
Further, Section 2 devotes to the review of some existing prominent direct estimators based on bivariate auxiliary information along with their MSE expressions. Section 3 proposes log type estimator based on bivariate auxiliary information under SRS. Evaluation of the proposed estimator using unit level real data taken from Sarndal et al. (2003) is done in Section 4. In Section 5, the conclusion is provided.

2. **Existing direct estimators for domain mean**

In this section, we present few existing direct estimators of domain mean under SRS using bivariate auxiliary information.

The direct mean per unit estimator is given by

$t_{m,a}^{d} = \bar{y}_a$

The variance of the estimator $t_{m,a}^{d}$ is given by

$V(t_{m,a}^{d}) = f_a \bar{Y}_a^2 C_{y_a}^2$

The direct ratio estimator based on bivariate auxiliary information is reported below as

$$t_{r,a}^{d} = \bar{y}_a \left(\frac{\bar{X}_a}{\bar{x}_a}\right)\left(\frac{\bar{Z}_a}{\bar{z}_a}\right)$$

The MSE of the estimator $t_{r,a}^{d}$ is given by

$MSE(t_{r,a}^{d}) = f_a \bar{Y}_a^2 (C_{y_a}^2 + C_{x_a}^2 + C_{z_a}^2 - 2\rho_{y_a x_a} C_{y_a} C_{x_a} - 2\rho_{y_a z_a} C_{y_a} C_{z_a} + 2\rho_{x_a z_z} C_{x_a} C_{z_a})$

The direct generalized ratio estimator based on bivariate auxiliary information is given by

$$t_{(j),a}^{d} = \bar{y}\left(\frac{a\bar{X}_a + b}{a\bar{x}_a + b}\right)\left(\frac{c\bar{Z}_a + d}{c\bar{z}_a + d}\right)$$

The MSE of the estimator $t_{(j),a}^{d}$ is given by

$$MSE\left(t_{(j),a}^{d}\right) = f_a\bar{Y}_a^2\left(C_{y_a}^2 + v_1^2C_{x_a}^2 + v_2^2C_{z_a}^2 - 2v_1\rho_{y_ax_a}C_{y_a}C_{x_a} - 2v_2\rho_{y_az_a}C_{y_a}C_{z_a} + 2v_1v_2\rho_{x_az_a}C_{x_a}C_{z_a}\right)$$

where $v_1 = \frac{a\bar{X}_a}{a\bar{X}_a + b}$ and $v_2 = \frac{c\bar{Z}_a}{c\bar{Z}_a + d}$; a, b, c, and d are either real values or function of known population parameters such as coefficient of kurtosis, standard deviation, coefficient of variation of auxiliary variables x and z, and coefficient of correlation of study and auxiliary variables.

Table 1. Few members of the generalized class of direct estimators $t_{(j),a}^{d}$

Members of the estimator $t_{(j),a}^{d}$	Values of			
$j = 1, 2, ..., 10$	a	b	c	d
$t_{1(a)}^{d} = \bar{y}_a\left[\frac{\bar{X}_a + \beta_2(x_a)}{\bar{x}_a + \beta_2(x_a)}\right]\left[\frac{\bar{Z}_a + \beta_2(z_a)}{\bar{z}_a + \beta_2(z_a)}\right]$	1	$\beta_2(x_a)$	1	$\beta_2(z_a)$
$t_{2(a)}^{d} = \bar{y}_a\left[\frac{\bar{X}_a + C_{x_a}}{\bar{x}_a + C_{x_a}}\right]\left[\frac{\bar{Z}_a + C_{z_a}}{\bar{z}_a + C_{z_a}}\right]$	1	C_{x_a}	1	C_{z_a}
$t_{3(a)}^{d} = \bar{y}_a\left[\frac{\beta_2(x_a)\bar{X}_a + C_{x_a}}{\beta_2(x_a)\bar{x}_a + C_{x_a}}\right]\left[\frac{\beta_2(z_a)\bar{Z}_a + C_{z_a}}{\beta_2(z_a)\bar{z}_a + C_{z_a}}\right]$	$\beta_2(x_a)$	C_{x_a}	$\beta_2(z_a)$	C_{z_a}
$t_{4(a)}^{d} = \bar{y}_a\left[\frac{C_{x_a}\bar{X}_a + \beta_2(x_a)}{C_{x_a}\bar{x}_a + \beta_2(x_a)}\right]\left[\frac{C_{z_a}\bar{Z}_a + \beta_2(z_a)}{C_{z_a}\bar{z}_a + \beta_2(z_a)}\right]$	C_{x_a}	$\beta_2(x_a)$	C_{z_a}	$\beta_2(z_a)$
$t_{5(a)}^{a} = \bar{y}_a\left[\frac{\bar{X}_a + \rho_{y_ax_a}}{\bar{x}_a + \rho_{y_ax_a}}\right]\left[\frac{\bar{Z}_a + \rho_{y_az_a}}{\bar{z}_a + \rho_{y_az_a}}\right]$	1	$\rho_{y_ax_a}$	1	$\rho_{y_az_a}$
$t_{6(a)}^{d} = \bar{y}_a\left[\frac{C_{x_a}\bar{X}_a + \rho_{y_ax_a}}{C_{x_a}\bar{x}_a + \rho_{y_ax_a}}\right]\left[\frac{C_{x_a}\bar{Z}_a + \rho_{y_az_a}}{C_{x_a}\bar{z}_a + \rho_{y_az_a}}\right]$	C_{x_a}	$\rho_{y_ax_a}$	C_{z_a}	$\rho_{y_az_a}$
$t_{7(a)}^{d} = \bar{y}_a\left[\frac{\rho_{y_ax_a}\bar{X}_a + C_{x_a}}{\rho_{y_ax_a}\bar{x}_a + C_{x_a}}\right]\left[\frac{\rho_{y_az_a}\bar{Z}_a + C_{z_a}}{\rho_{y_az_a}\bar{z}_a + C_{z_a}}\right]$	$\rho_{y_ax_a}$	C_{x_a}	$\rho_{y_az_a}$	C_{z_a}
$t_{8(a)}^{a} = \bar{y}_a\left[\frac{\beta_2(x_a)\bar{X}_a + \rho_{y_ax_a}}{\beta_2(x_a)\bar{x}_a + \rho_{y_ax_a}}\right]\left[\frac{\beta_2(z_a)\bar{Z}_a + \rho_{y_az_a}}{\beta_2(z_a)\bar{z}_a + \rho_{y_az_a}}\right]$	$\beta_2(x_a)$	$\rho_{y_ax_a}$	$\beta_2(z_a)$	$\rho_{y_az_a}$
$t_{9(a)}^{d} = \bar{y}_a\left[\frac{\rho_{y_ax_a}\bar{X}_a + \beta_2(x_a)}{\rho_{y_ax_a}\bar{x}_a + \beta_2(x_a)}\right]\left[\frac{\rho_{y_az_a}\bar{Z}_a + \beta_2(z_a)}{\rho_{y_az_a}\bar{z}_a + \beta_2(z_a)}\right]$	$\rho_{y_ax_a}$	$\beta_2(x_a)$	$\rho_{y_az_a}$	$\beta_2(z_a)$
$t_{10(a)}^{d} = \bar{y}_a\left[\frac{S_{x_a}\bar{X}_a + \beta_2(x_a)}{S_{x_a}\bar{x}_a + \beta_2(x_a)}\right]\left[\frac{S_{z_a}\bar{Z}_a + \beta_2(z_a)}{S_{x_a}\bar{z}_a + \beta_2(z_a)}\right]$	S_{x_a}	$\beta_2(x_a)$	S_{z_a}	$\beta_2(z_a)$

3. Proposed estimator

Motivated by the studies of Bhushan and Kumar (2022), we propose a logarithmic type direct estimators for domain mean using bivariate auxiliary information under SRS. The proposed direct estimator based on bivariate auxiliary information is given by

$$t_{p,a}^{d} = \bar{y}_a \left[1+\log\left(\frac{\overline{X}_a}{\bar{x}_a}\right)\right]^{\lambda_a}\left[1+\log\left(\frac{\overline{Z}_a}{z_a}\right)\right]^{\delta_a}$$

where λ_a and δ_a are constants.

To find the MSE and minimum MSE of the proposed direct estimator $t_{p,a}^{d}$, we utilize the notations provided in the earlier section and rewrite the proposed direct estimator $t_{p,a}^{d}$ as

$$t_{p,a}^{d} = \overline{Y}_a\left(1+e_o\right)\left[1+\log\left(\frac{\overline{X}_a}{\overline{X}_a\left(1+e_1\right)}\right)\right]^{\lambda_a}\left[1+\log\left(\frac{\overline{Z}_a}{\overline{Z}_a\left(1+e_2\right)}\right)\right]^{\delta_a}$$

After simplifying and subtracting $\overline{Y}_a$ on both sides, we get

$$t_{p,a}^{d}-\overline{Y}_a = \overline{Y}_a\left(1+\lambda_a e_1+\delta_a e_2-\lambda_a\frac{e_1^2}{2}-\delta_a\frac{e_2^2}{2}+\lambda_a^2\frac{e_1^2}{2}+\delta_a^2\frac{e_2^2}{2}+\lambda_a\delta_a e_1 e_2\right)-\overline{Y}_a$$

Squaring and taking expectation on both sides, we get

$$M\left(t_{p,a}^{d}\right) = f_a\overline{Y}_a^2\left(C_{y_a}^2+\lambda_a^2C_{x_a}^2+\delta_a^2C_{z_a}^2+2\lambda_a\rho_{y_ax_a}C_{y_a}C_{x_a}+2\delta_a\rho_{y_az_a}C_{y_a}C_{z_a}+2\lambda_a\delta_a\rho_{x_az_a}C_{x_a}C_{z_a}\right)$$

Minimization the above equation with respect to λ_a and δ_a provides the optimum value of λ_a and δ_a as

$$\lambda_{a(opt)} = \left(\frac{C_{y_a}}{C_{x_a}}\right)\left(\frac{\rho_{y_az_a}\rho_{x_az_a}-\rho_{y_ax_a}}{1-\rho_{x_az_a}^2}\right)$$

$$\delta_{a(opt)} = \left(\frac{C_{y_a}}{C_{z_a}}\right)\left(\frac{\rho_{y_ax_a}\rho_{x_az_a}-\rho_{y_az_a}}{1-\rho_{x_az_a}^2}\right)$$

Putting the above optimum values of λ_a and δ_a in the $MSE(t_{p,a}^{d})$, we get

$$min.MSE\left(t_{p,a}^{d}\right) = f_a\overline{Y}_a^2C_{y_a}^2\left(1-R_{y_a.x_az_a}^2\right)$$

where $R_{y_a.x_az_a}^2$ is the multiple correlation coefficient.

4. Numerical study

In this section, an empirical study is presented using unit level real data of Sweden municipalities reported in Sarndal et al. (2003). The municipalities of Sweden are divided into eight domains having sizes 25, 48, 32, 38, 56, 41, 15, and 29, respectively. The description of the variables is given below.

Y: REV84 = Real estate values according to 1984 assessment (in millions of Kronor),

X: P75 = 1975 population (in thousands),

Z: ME84 = Number of municipal employees in 1984,

The domain parameters of the data set are reported in Table 2.

Using the domain parameters, we have tabulated MSE and PRE of the proposed direct and synthetic estimators with the help of following formula:

$$PRE\left(t_{m,a}^{d}, t_{*,a}^{d}\right) = \frac{MSE\left(t_{m,a}^{d}\right)}{MSE\left(t_{*,a}^{d}\right)}\times 100$$

The results of the direct estimators are reported in Tables 2-3 in the form of MSE and PRE, respectively. From the table 3, it can be seen that the proposed direct estimator $t_{p,a}^{d}$ attains the least MSE and maximum PRE among the existing direct estimators such as direct mean estimator $t_{m,a}^{d}$, direct ratio estimator $t_{r,a}^{d}$, and direct generalized estimator $t_{(j),a}^{d}$ where $j=1,2,...,10$. Thus, the proposed bivariate auxiliary information based direct estimator outperforms the existing bivariate auxiliary information based direct estimators.

Table 2: Parameters of different domains

Domains	N_a	n_a	$\overline{Y}_a$	$\overline{X}_a$	$\overline{Z}_a$	S_{y_a}	S_{x_a}	S_{z_a}	$\rho_{y_a x_a}$	$\rho_{y_a z_a}$	$\rho_{x_a z_a}$
1	25	5	6413.32	59.52	4076.36	11317.06	128.70	8696.66	0.99	0.99	0.99
2	48	10	2971.10	29.17	1658.71	3334.66	35.05	2145.20	0.96	0.97	0.99
3	32	6	2498.75	23.94	1317.03	2040.72	20.91	1410.55	0.95	0.93	0.95
4	38	8	2915.53	30.63	1937.71	3094.46	41.49	3998.27	0.98	0.95	0.97
5	56	11	3046.46	28.71	1950.39	5278.27	59.71	6227.87	0.98	0.97	0.99
6	41	8	2175.32	20.98	1099.76	1693.82	17.35	1010.17	0.98	0.98	0.99
7	15	3	3648.47	26.60	1533.87	2410.56	24.12	1482.13	0.84	0.84	0.99
8	29	6	2269.10	17.14	1056.24	2784.95	20.15	1372.38	0.81	0.82	0.99

Table 3: MSE of different direct estimators

Domains Estimators	1	2	3	4	5	6	7	8
$t^d_{m,a}$	20492138	880332	563945	944969	2035240	288652	1549549	1025214
$t^d_{r,a}$	42812255	1438946	1163415	4806880	8753509	472259	6607974	1830916
$t^d_{(1),a}$	25176689	1076688	738681	2816776	3523047	290308	5335050	1295005
$t^d_{(2),a}$	40277837	1345180	1104386	4576514	8068158	442723	6327143	1679738
$t^d_{(3),a}$	42692026	1421444	1156826	4795371	8736694	468009	6548756	1801690
$t^d_{(4),a}$	32706985	1126304	703653	3156198	4939896	267619	5234393	1352984
$t^d_{(5),a}$	41615898	1362784	1099889	4639670	8419962	437830	6348097	1724329
$t^d_{(6),a}$	42251904	1375162	1091529	4683146	8590816	431251	6322836	1739561
$t^d_{(7),a}$	40261713	1342143	1101352	4572195	8054260	442021	6275343	1647572
$t^d_{(8),a}$	42756944	1424856	1156307	4798609	8745669	467265	6553310	1810766
$t^d_{(9),a}$	25101697	1067074	724414	2794093	3483077	287342	5144823	1215048
$t^d_{(10),a}$	42604172	1426081	1133052	4721600	8504139	457224	6544049	1793597
$t^d_{p,a}$	270936	55736	51341	26934	85259	12866	461387	336737

Table 4: PRE of different direct estimators

Domains Estimators	1	2	3	4	5	6	7	8
$t^d_{m,a}$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
$t^d_{r,a}$	47.87	61.18	48.47	19.66	23.25	61.12	23.45	55.99
$t^d_{(1),a}$	81.39	81.76	76.34	33.55	57.77	99.43	29.04	79.17
$t^d_{(2),a}$	50.88	65.44	51.06	20.65	25.23	65.20	24.49	61.03
$t^d_{(3),a}$	47.99	61.93	48.75	19.71	23.30	61.68	23.66	56.90
$t^d_{(4),a}$	62.65	78.16	80.15	29.94	41.20	107.86	29.60	75.77
$t^d_{(5),a}$	49.24	64.60	51.27	20.37	24.17	65.93	24.41	59.46
$t^d_{(6),a}$	48.49	64.02	51.67	20.18	23.70	66.93	24.51	58.94
$t^d_{(7),a}$	50.89	65.59	51.21	20.67	25.27	65.30	24.69	62.23
$t^d_{(8),a}$	47.93	61.78	48.77	19.69	23.27	61.77	23.65	56.62
$t^d_{(9),a}$	81.64	82.50	77.85	33.82	58.43	100.46	30.12	84.38
$t^d_{(10),a}$	48.10	61.73	49.77	20.01	23.93	63.13	23.68	57.16
$t^d_{p,a}$	**7563.44**	**1579.48**	**1098.44**	**3508.42**	**2387.12**	**2243.47**	**335.85**	**304.46**

From the empirical results reported in Tables 2-3, we can observe that the proposed bivariate logarithmic estimator dominates the existing mean estimator, bivariate ratio estimator, and bivariate generalized estimator in form of MSE and PRE, respectively in each domain.

5. Conclusion

In this article, we proposed a bivariate auxiliary information based logarithmic type direct estimator under SRS. The MSE expression of the proposed estimator is obtained approximately to the order one. The performance of the proposed estimator is compared with the existing estimators using an empirical study based on a real data set of Sweden municipalities. The results of the existing and proposed estimators are reported in Tables 3-4 in the form of MSE and PRE. The proposed estimator obtained least MSE and maximum PRE as compare to the existing estimators. Therefore, the proposed estimator is justified and can be recommended for the estimation of small domains.

References

Bhushan, S. and Kumar, A. (2022). New efficient logarithmic estimators using multi auxiliary information under ranked set sampling. Concurrency and Computation: Practice and Experience, 34(27), e7337.

Gonzalez, M.E. (1973). Use and evaluation of synthetic estimates. Proceedings of the social statistics section, American Statistical Association, 33-36.

Khare, B.B. and Ashutosh (2018). Simulation study of the generalized synthetic estimator for domain mean in the sample survey. International Journal of Tomography & Statistics, 31(3), 87-100.

Rai, P.K. and Pandey, K.K. (2013). Synthetic estimators using auxiliary information in small domains. Statistics in Transition- New Series, 14(1), 31-44.

Rao, J.N.K. and Molina, I. (2003). Small Area Estimation, John Wiley & Sons, Inc. Hoboken, New Jersey.

Sarndal, C.E., Swensson, B. and Wretman, J. (2003). Model Assisted Survey sampling, Springer-Verlog New York.

Tikkiwal, G.C. and Ghiya, A. (2000). A generalized class of synthetic estimators with application to crop acreage estimation for small domains. Biometrical Journal, 42, 865-876.