Journal of
Applied Sciences

Using Meta-analysis for Data Enrichment-optimal Families of Estimation Strategies

[1]Ashok Sahai, [1]M. Mahibbur Rahman and [2]Rameshwar P. Jaju
[1]Department of Statistics and Demography,
[2]Department of Computer Science,
University of Swaziland, Kwaluseni, P. Bag 4, Swaziland, Africa

Abstract: Long-term data-sets of good quality are invaluable in certain research investigations. In some situation, they are so expensive of time and money, that they are seldom available. It has been noted that the problems of varying data quality, of missing information and that of diversity of data might possibly be dealt with by using meta-data. The key point is to treat relevant good quality meta-data as auxiliary/ancillary information, the use of which is gained through the proposed optimal simple estimation strategies. These strategies are motivated by mixing-type estimators, the desired optimality of these estimators is achieved through optimal manipulation of two design-parameters therein set to control first and second order of large sample approximations to their standard errors. A sensitivity analysis is then used to discover, empirically, the robust estimation strategy from amongst the proposed alternatives.

Key words: Meta-analysis, large sample approximation, sensitivity analysis

INTRODUCTION

Meta-analysis can be a useful tool in several research experiments where certain data are directly effected by natural events. The examples could include environmental research, forestry inventory control, sea food harvesting management, resource intensive or very costly natural experiments and other similar problems. For the purpose of our discussion, we have used the example of environmental research.

Many investigators, including Cuthbertson[1], have highlighted the value of meta-data and meta-analysis in environmental studies. For example, it is worth quoting from Hedges and Olkin[2]. "Finally, some notice might be in order regarding the usefulness of meta-analysis in environmental research. Using meta-analysis a researcher can ascertain which issues are likely to reward additional inquiry. In addition, at the peer-review level, which issues are likely to reward additional inquiry. In addition, at the peer-review level, meta-analysis can be used to synthesize apparently conflicting results into a unified corpus of knowledge". Meta-analysis basically deals with techniques of synthesis of statistical inferences from two different/independent studies, possibly different spatially and/or at two different points of time. One of the currently investigated areas, receiving intensive attention of researchers in this field, is the handling of studies which might be dependent: possibly also with missing data[3,4].

From meta-data perspective, we would almost always have, at the meta-level, the temporal and spatial gaps in the data. Because of fast-changing environmental system-parameters in the present decade (by comparison with the years of, say, the preceding decade), time-series methods might not be adequate if we wish to have synthetic data for relevant system variables. On the contrary, using the preceding decade's meta-data might well capture this fast-changing pattern more adeptly than the use of time series methods. It will, therefore, do better if treated via the use of ancillary/auxiliary variables setup, to capture this rather-faster-changing pattern more adeptly than via the time-series methods. In addition, there might be many spatial gaps where the underdeveloped and developing countries have insufficient resources, not enough to monitor many of the environmental system-variables, for some variables, data may not be available at all. To fill these gaps, small-scale surveys may be conducted by international agencies. However. Some of the variables may continue to be missing due to practical reasons, such as the lack of suitable infrastructure for the experimental work, etc.

Therefore, same strategy of using information on the ancillary/auxiliary variables would be appropriate for those missing variables. It is necessary to devise an estimation strategy which would exploit the high correlation (positive or negative) between the study variables and ancillary/ auxiliary variables. There are two

Corresponding Author: Rameshwar P. Jaju, Department of Computer Science, University of Swaziland,
Kwaluseni, P. Bag 4, Swaziland, Africa

main facts to be noted in this context: firstly, the usable data on such auxiliary/ancillary variables may not be extensive. The available data on such variables might have high dispersion temporally and/or spatially. Therefore it might not be advisable to use more than say, ten to twenty years' data due to rapidly changing environmental system. Similarly in the other case, due to high spatial variability of data it could be equally undesirable to use data on the relevant auxiliary variable for more than ten or possibly less number of countries or sites with similar environmental zone where such a data becomes available. Secondly, due to system-dynamics of the environs and its rapidly changing patterns (spatially and temporally), the dispersion of the relevant meta-data on such an auxiliary/ancillary variables would be relatively very high. This has to be tackled statistically in as much as the variances of sub-groups (spatial/temporal) of data on such variables should not be highly significantly different to be synthesized in meta-data setup.

The data on the environmental variables are usually collected using small samples. Otherwise the data become very expensive. In some instances, if the study continues for longer period, the information on certain environmental variable becomes more expensive. Specially in the case of developing countries, it is very common to face the problem of missing information (spatial and/or temporal) on environmental variables due to limited availability of resources to monitor those. To overcome these difficulties, we might be better off is we tackle the gaps in the data using meta-data setup by searching for highly correlated (but relatively dispersed) variables having auxiliary/ancillary status of information for the study variable.

EXAMPLE APPLICATION

To illustrate, survey results for 19 landslide-damaged sites are reported by Reddy and Singh[5]. Eleven of these sites were in Oak-zone and 8 were in Pine or mixed Pine zones. Vegetation analysis was carried out for 10 one meter-squared quadrants: Biomass of the herb layer at the peak growth stage (i.e. in last week of august) was determined for 5 soil monoliths (25x25x30 cm (deep)) excavated randomly. Soil samples were collected in triplicate, from each site from 0-10 cm soil depths. Statistical analysis was done via the analysis of variance and non-linear regressions. We note that all the data were collected using small samples. Here the variable (age of site in years), say X, was found to be a good auxiliary variable. The relative dispersion for X was high: $C_X = 0.98$ for Pine-zone and $C_X = 1.09$ for Oak-zone: It was almost

double of the relative dispersion of one of the important study variable, namely, A:P (Annual to Perennial (Herb-species) Ratio), say Y, with $C_Y = 0.48$ for the Pine-zone and $C_Y = 0.57$ for the Oak-zone. The corresponding values of C_Y for the other study variables for these two zones were (0.18, 0.28), (0.10, 0.19), (0.50, 0.42) and (0.62, 0.61); when Y ≡ Species, Annuals, Perennial and Cover, respectively and were as high as 0.81 and 0.90, when Y ≡ Density. Reddy and Singh[5] reported many couples of variables with a significantly high correlation: ± 0.6 to as high as 0.972. For the one relevant to our example, the coefficient of correlation between the study variables Y (i.e. A:P Ratio) and the auxiliary variables X (i.e. age of site) are fairly significant as reported in the paper: -0.718 for Pine-zone and –0.420 for Oak-zone.

Thus in the context of environmental database management in the meta-data setup and in the light of the aforesaid facts, we need to devise (for small sample) an estimation strategy for transporting the ancillary information, contained in as auxiliary variable (X). This variable X would be having a significant correlation with the study variable (Y); whereas, X-data might well be rather much more dispersed than the Y-data.

THE ESTIMATORS

The estimation strategies proposed in this study are motivated by the mixing-type estimation of Vos[6]. Sahai[7] used auxiliary information efficiently in his mixing type estimators. The same goal of efficiency is addressed here using auxiliary information (when the sample is small, i.e. n<30) for the problem of estimating the population average of the study variable. Subsequently, it can be used in aggregation or disaggregation of meta-data. In practice, it is possible to get hold of such an auxiliary variable with population mean (e.g. population average of age (in years) of the Oak-zone/Pine-zone sites in the proceeding illustration) which could either be known or could be know using past information.

With the above motivation, the following families of estimates are proposed. Each of these two families of estimates consists of two non-stochastic design parameters, namely α and β:

$$\overline{y}_{\alpha\beta}(1) = \overline{y}[\{\alpha\overline{x} + (1-\alpha)\overline{X}\}/\{\beta\overline{x} + (1-\beta)\overline{X}\}] \quad (1)$$

and

$$\overline{y}_{\alpha\beta}(2) = \overline{y}[\{\alpha\overline{x} + (1-\alpha)\overline{X}\}/\{\beta\overline{x} + (1-\beta)\overline{X}\}]^2 \quad (2)$$

Where, $\overline{x}$ and $\overline{y}$ are the sample means of the auxiliary and study variables, respectively and $\overline{X}$ is the population mean of auxiliary variable. The choice of the

appropriate values for the two design-parameters is governed by their roles to minimize the first order approximation (upto the terms of $O(n^{-1})$ and the second order approximation (upto terms $O(n^{-2})$) to the standard error of the relevant estimators. Further, the use of estimators in these families is highly recommended, provided the absolute value of the quantity within square brackets of Eq. 1 and 2 (i.e., the perturbators for $\overline{y}$) stay away from unity on its either side by upto 30 percent of the quantity $(1+2\rho^2)$ where, ρ is the coefficient of correlation between the two variables X and Y. We may note that the term $(1+2\rho^2)$ is the leading multiplier of the second order approximation of the mean square error Eq. 10 and 11. The simulation study reported that only in 3% to 9% of the cases (depending on ρ values) the perturbations stay out of these bounds.

It may be noted that the mixing estimators of Vos[6] are:

$$\overline{y}_M(1) = \alpha\overline{y} + (1-\alpha)\overline{y}_P \qquad (3)$$

and

$$\overline{y}_M(2) = \alpha\overline{y} + (1-\alpha)\overline{y}_Q \qquad (4)$$

wherein $\overline{y}_P = (\overline{y}\,\overline{x})/\overline{X}$ and $\overline{y}_Q = (\overline{y}/\overline{x})\overline{X}$ are the well-known product and quotient estimators, respectively and α is the design-parameter that minimizes the first order approximation ($O(n^{-1})$) to the standard error of the estimator.

On the other hand structurally, the proposed families of estimators happen to be a result of marrying the perspective of the one-parameter, family of the estimators, $\overline{y}_{sa}$[7] and the one with that of $\overline{y}_{sr}$ [8]. These are, respectively:

$$\overline{y}_{sa} = \overline{y}[\{\alpha\overline{X} + (1-\alpha)\overline{x}\}/\{\alpha\overline{x} + (1-\alpha)\overline{X}\}] \qquad (5)$$

$$\overline{y}_{sr} = \overline{y}(\overline{X}/\overline{x})^\alpha \qquad (6)$$

where, in both cases the value of the design-paramater, α is chosen similarly as in the case of the estimators of Vos[6], so that it minimizes the first order approximation ($O(n^{-1})$) to the standard error of the respective estimators. It is worth noting that the proposed families of estimators inherit the perspective of a version of $\overline{y}_{sr}$ [8] and that a fractional α is not computationally favorable to numerical exponential approximation. Another such estimator with a similar perspective $\overline{y}_{re}$ has been given by Reddy[9]:

$$\overline{y}_{re} = (\overline{y}\,\overline{X})/\{\beta\overline{x} + (1-\beta)\overline{X}\} \qquad (7)$$

Note that the above estimator is embedded in the proposed family of estimators $\overline{y}_{\alpha\beta}$ (1) with $\alpha = 0$, i.e. one-parameter sub-family without the facility of exploitation of a second degree of freedom. Then, the only parameter β controls both first and second order of approximations to the mean square error.

Using the result of bivariate normal population[10], we get the following expressions for the first and second order of approximations to the Mean Square Error (MSE) of the relevant estimators ($MSE_1(.)$ and $MSE_2(.)$), respectively

$$MSE_1(\overline{y}_{\alpha\beta}(1)) = (\overline{Y}^2/n)\left\{C_y^2 + (\alpha-\beta)^2 C_x^2 \right. \\ \left. + 2(\alpha-\beta)kC_x^2\right\} \qquad (8)$$

$$MSE_1(\overline{y}_{\alpha\beta}(2)) = (\overline{Y}^2/n)\left\{C_y^2 + 4(\alpha-\beta)^2 C_x^2 \right. \\ \left. + 4(\alpha-\beta)kC_x^2\right\} \qquad (9)$$

$$MSE_2(\overline{y}_{\alpha\beta}(1)) = MSE_1(\overline{y}_{\alpha\beta}(1)) \\ + (C_x^4\overline{Y}^2/n^2)[6k\beta(\alpha-\beta)(3\beta-2\alpha) \\ + (1+2\rho^2)(k^2/\rho^2)(\alpha-\beta)(\alpha-3\beta) \\ + 9\beta^2(\alpha-\beta)^2] \qquad (10)$$

and

$$MSE_2(\overline{y}_{\alpha\beta}(2)) = MSE_1(\overline{y}_{\alpha\beta}(2)) + (C_x^4\overline{Y}^2/n^2) \\ [12k(\alpha-\beta)\{2(\alpha-\beta)(\alpha-3\beta) - \beta(\alpha-2\beta)\} \\ + 2(1+2\rho^2)(k^2/\rho^2)(\alpha-\beta)\{3(\alpha-\beta) - 2\beta\} \\ + 3(\alpha-\beta)^2\{(\alpha-2\beta)^2 - 8\beta(\alpha-2\beta)\}] \qquad (11)$$

where, $k = \rho(C_y/C_x)$. By minimizing $MSE_1(.)$ and $MSE_2(.)$ in Eqs. 8 to 11, we get:

$$\alpha = \beta - k \quad ; \beta = (4\rho^2-1)k/(3\rho^2) \text{ for } \overline{y}_{\alpha\beta}(1) \qquad (12)$$

$$\alpha = \beta - k/2 \quad ; \beta = (13\rho^2-4)k/(12\rho^2) \text{ for } \overline{y}_{\alpha\beta}(2) \qquad (13)$$

To use these optimal values of design-parameters, we are constrained by the lack of knowledge of the values of ρ and k, where k is a function of ρ and C_y/C_x (Eq.12 and 13). However, we use the easily available close guesses on ρ and C_y/C_x via the past data or long association with the experimental setup in order to obtain a value for k.

To study the sensitivity (robustness) of the estimators in the proposed families to the relative errors in guessing k as also to study their relative efficiencies in

Table 1: Relative efficiency for different ratio-product estimators

n	REG(k)%	$\rho = 0.1$				$\rho = -0.1$			
		$\overline{y}_M(1)$	$\overline{y}_{\alpha\beta}(1)$	$\overline{y}_{\alpha\beta}(2)$	$\overline{y}_{re}$	$\overline{y}_M(1)$	$\overline{y}_{\alpha\beta}(1)$	$\overline{y}_{\alpha\beta}(2)$	$\overline{y}_{re}$
5	-20	100.78	95.02	95.04	100.53	100.66	95.41	95.43	100.42
	-15	100.78	94.54	94.55	100.49	100.65	94.54	94.54	100.37
	-10	100.77	94.50	94.59	100.45	100.64	94.20	94.20	100.33
	-5	100.76	94.45	94.44	100.40	100.62	94.49	94.50	100.27
	0	100.74	93.84	93.89	100.34	100.59	94.46	94.47	100.20
	5	100.71	93.13	93.12	100.27	100.56	94.40	94.21	100.13
	10	100.68	92.13	92.13	100.20	100.52	94.65	94.59	100.05
	15	100.64	92.64	92.74	100.11	100.48	94.65	94.64	99.96
	20	100.60	92.27	92.33	100.02	100.42	94.58	94.62	99.86
10	-20	100.53	94.32	94.23	100.42	101.04	94.50	94.52	100.91
	-15	100.52	94.77	94.71	100.39	101.06	94.38	94.35	100.91
	-10	100.50	93.69	93.75	100.35	101.07	94.04	94.06	100.91
	-5	100.48	93.38	93.41	100.31	101.08	94.43	93.45	100.89
	0	100.45	94.01	93.91	100.26	101.08	94.73	93.78	100.87
	5	100.41	93.91	93.86	100.21	101.07	94.76	93.79	100.85
	10	100.37	95.61	95.54	100.15	101.06	94.72	92.62	100.82
	15	100.33	95.52	95.40	100.08	101.04	94.20	93.26	100.78
	20	100.27	94.64	94.68	100.01	101.02	94.26	94.28	100.73
20	-20	100.84	93.76	93.82	100.79	101.15	94.08	94.11	101.09
	-15	100.85	94.59	94.57	100.79	101.18	94.14	93.13	101.10
	-10	100.86	95.10	95.12	100.79	101.20	94.21	94.22	101.12
	-5	100.85	95.15	95.19	100.78	101.22	94.78	93.62	101.12
	0	100.84	96.13	96.17	100.76	101.23	94.98	94.65	101.13
	5	100.83	94.67	94.68	100.74	101.24	94.47	94.54	101.12
	10	100.81	95.15	95.18	100.71	101.24	94.86	94.91	101.11
	15	100.79	95.40	95.44	100.68	101.23	94.71	95.74	101.09
	20	100.76	95.76	95.61	100.64	101.22	94.28	96.30	101.07

Table 2: Relative efficiency for different ratio-product estimators

n	REG (k)%	$\rho = 0.3$				$\rho = -0.3$			
		$\overline{y}_M(1)$	$\overline{y}_{\alpha\beta}(1)$	$\overline{y}_{\alpha\beta}(2)$	$\overline{y}_{re}$	$\overline{y}_M(1)$	$\overline{y}_{\alpha\beta}(1)$	$\overline{y}_{\alpha\beta}(2)$	$\overline{y}_{re}$
5	-20	106.93	107.01	107.01	104.55	106.62	106.50	106.55	104.52
	-15	106.85	106.96	107.11	104.15	106.54	106.52	106.53	104.17
	-10	106.70	106.93	106.99	103.66	106.40	106.38	106.48	103.75
	-5	106.49	106.58	106.67	103.07	106.21	106.10	105.93	103.24
	0	106.22	106.39	106.46	102.38	105.95	105.76	105.79	102.65
	5	105.89	106.00	106.08	101.60	105.64	105.40	105.55	101.97
	10	105.50	105.61	105.69	100.71	105.27	105.16	105.20	101.21
	15	105.05	105.20	105.19	99.71	104.84	104.63	104.66	100.36
	20	104.54	104.57	104.66	98.60	104.36	104.29	104.39	99.41
10	-20	109.59	110.13	110.04	108.36	108.87	109.86	109.91	107.68
	-15	109.72	110.10	110.13	108.35	108.98	110.00	109.99	107.62
	-10	109.79	110.02	110.09	108.26	109.02	110.08	110.13	107.49
	-5	109.80	110.14	110.16	108.10	109.00	110.11	110.07	107.29
	0	109.75	110.01	110.09	107.86	108.93	109.95	109.87	107.01
	5	109.64	109.96	110.04	107.55	108.79	109.78	109.82	106.65
	10	109.46	109.83	109.87	107.17	108.60	109.57	109.64	106.22
	15	109.23	109.55	109.57	106.71	108.35	109.29	109.33	105.71
	20	108.93	109.20	109.25	106.18	108.04	108.95	108.98	105.12
20	-20	108.85	109.27	109.29	108.34	109.83	110.80	110.82	109.22
	-15	109.00	109.47	109.47	108.42	110.02	110.86	110.91	109.34
	-10	109.09	109.45	109.49	108.45	110.16	111.02	111.03	109.39
	-5	109.13	109.50	109.53	108.41	110.24	111.06	111.08	109.39
	0	109.12	109.49	109.50	108.32	110.26	111.04	111.05	109.32
	5	109.05	109.38	109.40	108.16	110.23	110.98	111.01	109.19
	10	108.93	109.27	109.29	107.94	110.14	110.86	110.80	108.99
	15	108.75	109.08	109.10	107.67	109.99	110.61	110.62	108.73
	20	108.52	108.74	109.77	107.33	109.79	110.39	110.40	108.41

Table 3: Relative efficiency for different ratio-product estimators

n	REG (k)%	$\bar{y}_M(1)$	$\bar{y}_{\alpha\beta}(1)$	$\bar{y}_{\alpha\beta}(2)$	$\bar{y}_{re}$	$\bar{y}_M(1)$	$\bar{y}_{\alpha\beta}(1)$	$\bar{y}_{\alpha\beta}(2)$	$\bar{y}_{re}$
		$\rho = 0.5$				$\rho = -0.5$			
5	-20	122.16	121.83	121.81	102.92	125.87	124.67	124.96	118.99
	-15	121.73	121.19	121.23	99.28	125.69	124.97	124.96	118.46
	-10	121.07	121.23	121.40	99.27	125.25	125.98	126.27	117.84
	-5	120.17	122.09	122.04	90.94	124.57	127.39	127.68	117.13
	0	119.05	123.65	123.75	86.34	123.64	128.53	128.36	116.34
	5	117.72	124.87	124.77	81.54	122.47	130.64	130.84	115.47
	10	116.20	125.89	126.01	76.58	121.09	132.35	131.65	114.53
	15	114.49	127.09	126.91	71.55	119.50	133.58	133.86	113.52
	20	112.62	127.44	127.62	66.51	117.72	133.79	134.22	112.46
10	-20	127.61	120.85	127.73	124.72	128.78	128.69	128.79	126.05
	-15	127.79	120.73	127.83	124.54	128.98	129.25	129.29	126.03
	-10	127.73	120.93	128.02	124.05	128.93	129.12	129.24	125.73
	-5	127.44	120.88	127.99	123.26	128.63	129.66	129.72	125.15
	0	126.91	120.25	128.08	122.15	128.09	129.83	129.82	124.27
	5	126.15	129.31	128.37	120.71	127.30	129.87	129.97	123.08
	10	125.17	129.74	127.96	118.92	126.29	130.09	130.14	121.57
	15	123.99	129.64	127.87	116.78	125.06	131.24	131.40	119.73
	20	122.60	128.46	128.71	114.29	123.62	131.08	130.90	117.54
20	-20	128.85	128.83	128.83	127.41	130.03	130.30	130.30	128.64
	-15	129.05	129.11	129.12	127.70	130.48	130.60	130.60	128.97
	-10	129.23	129.20	129.22	127.76	130.69	130.72	130.74	129.06
	-5	129.18	129.13	129.16	127.57	130.67	130.66	130.68	128.89
	0	128.91	128.84	128.88	127.13	130.41	130.37	130.40	128.47
	5	128.42	128.73	128.78	126.44	129.92	129.85	129.92	127.78
	10	127.71	128.08	128.13	125.48	129.20	129.35	129.80	126.83
	15	126.79	127.12	127.19	124.27	128.26	129.11	129.18	125.63
	20	125.67	126.21	126.30	122.80	127.11	128.24	128.36	124.16

Table 4: Relative efficiency for different ratio-product estimators

n	REG(k)%	$\bar{y}_M(1)$	$\bar{y}_{\alpha\beta}(1)$	$\bar{y}_{\alpha\beta}(2)$	$\bar{y}_{re}$	$\bar{y}_M(1)$	$\bar{y}_{\alpha\beta}(1)$	$\bar{y}_{\alpha\beta}(2)$	$\bar{y}_{re}$
		$\rho = 0.7$				$\rho = -0.7$			
5	-20	154.83	178.77	178.77	125.26	146.92	179.28	178.27	100.32
	-15	45306	183.35	183.82	116.61	145.16	185.66	185.18	29.55
	-10	150.53	186.19	186.07	107.28	142.44	191.3	190.48	112.77
	-5	147.32	191.99	192.47	98.61	139.82	194.47	194.23	0.44
	0	143.53	194.32	191.95	87.87	136.34	199.25	198.92	2.12
	5	139.28	193.83	194.32	79.34	132.46	202.61	203.10	67.32
	10	134.57	195.85	195.65	69.22	128.25	202.77	203.25	32.84
	15	129.62	196.52	196.7	60.68	123.79	202.59	202.96	2.89
	20	124.48	199.1	199.79	52.81	119.15	203.15	204.02	7.75
10	-20	173.19	181.25	180.96	172.70	168.08	178.07	177.99	147.32
	-15	173.67	185.98	185.76	176.78	168.25	185.44	494.75	145.93
	-10	173.27	190.51	189.38	173.8	167.60	187389	187.78	144.25
	-5	172.00	194.12	194.18	172.59	166.15	190.56	190.84	142.22
	0	169.91	197.64	197.34	169.98	163.93	192.48	192.49	140.22
	5	167.04	200.27	199.98	165.8	161.02	193.80	193.56	137.91
	10	193.49	200.81	200.95	159.82	157.49	194.75	194.92	135.46
	15	159.34	200.30	199.85	151.78	153.42	195.62	195.96	132.90
	20	154.69	198.57	199.01	141.17	148.90	195.30	195.46	130.24
20	20	173.64	175.55	175.49	175.00	176.14	177.88	177.79	177.96
	-15	174.67	178.00	177.90	176.56	177.66	182.06	181.67	179.97
	-10	174.87	180.61	180.50	177.24	178.37	186.00	185.66	181.06
	-5	174.25	182.42	182.49	175.70	177.31	187.84	187.72	179.99
	0	172.81	182.42	182.49	175.70	177.31	187.84	187.72	179.99
	5	170.60	182.08	181.86	173.39	175.58	187.00	187.02	177.66
	10	167.68	180.84	180.63	170.05	173.09	187.74	187.24	174.03
	15	164.12	179.25	179.55	164.71	169.91	186.18	186.44	169.05
	20	160.01	177.73	177.69	160.42	166.13	183.96	184.10	162.71

Table 5: Relative efficiency for different ratio-product estimators

n	REG(k)%	$\rho = 0.9$				$\rho = -0.9$			
		$\overline{y}_M(1)$	$\overline{y}_{\alpha\beta}(1)$	$\overline{y}_{\alpha\beta}(2)$	$\overline{y}_{re}$	$\overline{y}_M(1)$	$\overline{y}_{\alpha\beta}(1)$	$\overline{y}_{\alpha\beta}(2)$	$\overline{y}_{re}$
5	-20	209.35	457.80	455.92	140.99	202.43	475.82	472.43	110.16
	-15	199.25	480.05	479.74	3.65	194.88	510.65	507.02	21.17
	-10	187.69	509.92	509.69	18.94	185.79	540.54	537.00	10.46
	-5	175.32	518.17	516.41	85.55	175.64	554.65	554.67	9.65
	0	162.69	524.02	522.34	32.49	164.91	568.73	567.28	3.81
	5	150.23	531.84	530.66	26.48	153.98	570.96	571.99	17.98
	10	138.25	530.52	532.83	1.65	143.17	571.31	573.42	8.89
	15	126.96	533.35	532.12	0.62	132.73	567.67	571.06	5.73
	20	116.45	538.77	537.11	2.12	122.79	562.97	566.94	0.36
10	-20	278.32	439.51	433.36	190.14	276.26	422.21	418.10	315.74
	-15	275.75	472.69	468.59	160.68	273.47	460.17	452.53	289.56
	-10	269.47	504.13	499.14	133.32	267.03	489.07	482.79	70.63
	-5	259.97	527.19	524.33	109.22	257.45	504.20	502.27	259.50
	0	247.94	538.04	537.82	88.73	245.40	511.25	512.12	324.13
	5	234.14	538.59	536.98	71.70	231.65	511.19	510.98	169.46
	10	219.31	534.83	534.60	57.77	261.92	507.12	507.81	151.23
	15	204.10	521.91	525.02	46.48	201.84	503.75	503.90	40.83
	20	189.02	508.82	508.76	37.36	186.91	492.92	494.89	8.02
20	-20	345.09	407.23	403.61	399.78	340.20	405.88	403.91	405.62
	-15	352.57	440.84	438.47	424.06	346.60	439.63	437.67	431.28
	-10	354.64	465.48	463.15	441.92	347.64	467.21	464.62	450.21
	-5	351.09	485.69	483.30	450.12	343.24	487.25	484.99	458.70
	0	342.27	487.53	487.70	445.79	333.79	491.69	492.30	453.26
	5	328.93	481.59	482.33	427.19	320.12	486.56	487.08	431.58
	10	312.16	470.65	472.51	394.28	303.27	476.15	477.80	393.07
	15	293.09	454.68	457.19	348.37	284.38	456.27	458.09	337.61
	20	272.83	433.72	437.38	290.72	264.48	434.37	437.72	255.30

order to discover the appropriate estimation strategy to be recommended in a practical situation, an extensive simulation study was carried out.

RESULTS AND DISCUSSION

For simulation study, we consider two independent normal populations with

$$\mu_Y = 6, \sigma_Y = 6 \text{ and } \mu_X = 6, \sigma_X = 12 \text{ so that } C_X = 2C_Y$$

We generate respective bivariate normal population using ten different values: ±0.1, ±0.3, ±0.5, ±0.7 and ±0.9 for three different sample sizes: $n = 5$, 10 and 20.

Also to discover the robustness of the estimators $\overline{y}_{\alpha\beta}(1)$ and $\overline{y}_{\alpha\beta}(2)$, we carry out sensitivity analysis using nine cases of the relative error in guessing k by using g, say -

$$REG(k) = [(g-k)/k]100\%$$

We have generated 5000 samples for each combination of n, ρ and REG(k) and have calculated the actual MSE of all the estimators mentioned (Eqs. 1 to 7). These results are summarized in

Table 1-5. The tabulated values represent the relative efficiencies of the estimators.

$$REF(.) = [MSE(\overline{y})/MSE(.)] \times 100\%$$

Table 1-5 are organized for different data of coefficient of correlation-(ρ), sample size (n) and REG (k) = 0, $\pm5\%$, $\pm10\%$, $\pm15\%$ and $\pm20\%$.

The results are encouraging for the proposed families of estimators. As expected, $\overline{y}_{sa}$ and $\overline{y}_{sr}$ din not perform so well as the estimators of the proposed families. The same happened to Vos'[6] $\overline{y}_M(2)$ as compared to his $\overline{y}_M(1)$. Hence we excluded these estimators while presenting these results.

The results also shows that:

- For $|\rho| \geq 0.5$, the proposed families of estimators are significantly more efficient as compared to $\overline{y}_M(1)$ and $\overline{y}_{re}$. Moreover, for larger values of ρ, these efficiencies become higher when the sample size is small.
- When the correlation is very low (say $|\rho| < 0.3$), the estimators $\overline{y}_M(1)$ and $\overline{y}_{re}$ performed better than the proposed estimators even though the improvement over $\overline{y}$ was rather insignificant.

- In the case to under-guess (i.e. when REG(k) is negative), $\overline{y}_{\alpha\beta}$ (1) is more favorable.
- For the case of the over-guess (i.e. when REG(k) is positive), $\overline{y}_{\alpha\beta}$ (2) performs better.

Therefore, in practice, it will be prudent to use the estimation strategy via the simple mixing:

$$\overline{y}_{\alpha\beta} = [\overline{y}_{\alpha\beta}(1) + \overline{y}_{\alpha\beta}(2)]/2,$$

since it would be unknown to us as to whether we are under-guessing of over-guessing.

REFERENCES

1. Cuthbertson, M., 1993. A database for environmental research programs. J. Environ. Manag., 37: 291-300.
2. Hedges, L.V. and I. Olkin, 1985. Statistical Methods for Meta-Analysis. Acad. Press, Orlando, Florida.
3. Hedges, L.V., 1990. The Future of Statistical Methods for Meta-Analysis. In: Straf, M. and K. Wachter (Eds.). The Future of Meta-Analysis, Russel Sage Foundation, New York, pp: 11-26.
4. Hedges, L.V., 1992. Meta-analysis. J. Edu. Stat., 17: 279-296.
5. Reddy, V.S. and J.S. Singh, 1993. Changes in vegetation and soil during succession following landslide disturbances in the central Himalaya. J. Environ. Manage., 39: 235-250.
6. Vos, J.W.E., 1980. Mixing of direct, ratio and product method estimations. Stat. Neerlandica, 34: 209-218.
7. Sahai, A., 1979. An efficient variant of the product and ration estimators. Stat. Neerlandica, 33: 27-35.
8. Srivastava, S.K., 1967. An estimator using auxiliary information in sample surveys. Calcutta Stat. Asso. Bull., 16: 121-132.
9. Reddy, V.N., 1974. On a transformed ratio method of estimation. Sankhya, C., 36: 59-70.
10. Sukhatme, P.V. and B.V. Sukhatme, 1970. Sampling Theory of Surveys with Applications. Iowa State Univ. Press, Am., Iowa.