Aligarh Journal of Statistics
Vol. 35(2015), 133-144

COMPROMISE ALLOCATION FOR COMBINED LINEAR REGRESSION ESTIMATES IN MULTIVARIATE STRATIFIED SAMPLING

Farha Naz, Ummatul Fatima, M. J. Ahsan and Q. M. Ali

ABSTRACT

In multivariate stratified sample surveys when auxiliary information is available it can be used to construct separate and combined ratio and regression estimates of the population means (see Khan et al. (2010)). This paper deals with the complex problem of obtaining a compromise allocation for constructing combined linear regression estimates of the population means of a multivariate stratified population when apart from the measurement cost there are also significant within stratum travelling costs resulting in a nonlinear cost constraint. The problem is formulated as a Multi-objective Integer Nonlinear Programming Problem (MINLPP) and a solution procedure is developed using Goal Programming Technique. The solution obtained is compared with some other allocations to show that the proposed procedure gives more precise result. The numerical results are obtained by using the optimization software LINGO.

1. INTRODUCTION

The problem of optimum allocation is a well known problem related to stratified sample surveys. Neyman (1934) gave the formula for working out optimum allocation for fixed total sample size. In a multivariate stratified population, where more than one characteristic are to be studied on each population unit, obtaining a unique set of sample size allocations to strata becomes a complex problem.

Various authors suggested different criteria, popularly known as compromise criteria, to work out an allocation that is optimum for all characteristics in one or other sense. Pioneering works in this regard were due to Dalenius (1957), Yates (1960), Kokan and Khan (1967), Chatterjee (1967) etc. Later on the earlier works are extended by many researchers like Ahsan and Khan (1977, 1982),

Bethel (1985, 1989), Chromy (1987), Jahan et al. (1994), Khan et al. (1997), Khan et al. (2003), Holmberg (2003), Najmussehar et al. (2005), Diaz Garcia and

Garay Tapia (2007), Kozak (2004, 2006), Ansari et al. (2009), Ansari et al. (2011) and many others.

In sample surveys a common practice is to use the auxiliary information if and when available to enhance the precision of the estimator of the population parameters under estimation. Ahsan and Khan (1982), Rao (1973), Dayal (1985), Ige and Tripathi (1987), Tripathi and Bahl (1991), Najmussehar and Bari (2002) and many others used auxiliary information to enhance the precision of the estimates in complex sample surveys. Most of the authors do not impose integer restrictions on the variables. In practice the resulting continuous solution is rounded off to the nearest integer value for practical applications. However, Khan et al. (1997) showed that the rounded off solution may prove infeasible or non-optimum. Khan et al. (2010) worked out compromise allocation for estimating the population means of a multivariate stratified population using combined ratio estimator. They formulated the problem as a multiobjective integer nonlinear programming problem and used Goal Programming Technique to obtain a solution.

In this paper the problem of determining a compromise allocation for a multivariate stratified population using auxiliary information has been formulated as a Multiobjective Integer Nonlinear Programming Problem (MINLPP) when combined regression estimators are used to estimate the population means as suggested by Khan et al. (2010). Techniques to solve a MINLPP are available in Multiobjective Programming literature. For example, Boer and Hendrix (2000), Hendrix et al. (2001), Chinchuluun and Pardalos(2007), Pardalos et al. (1995) and Zopounidis and Pardalos (2010) etc. Some recent related work can be seen in Floudas and Pardalos (2009). Since a number of optimization softwares are now available, one can use a suitable software to obtain a solution. Using an optimality criterion the authors used Goal Programming Technique to construct an Integer Nonlinear Programming Problem (INLPP) equivalent to the formulated MINLPP and obtained an integer solution directly by the optimization software LINGO (2010) as used in Khan et al. (2010). A numerical example is also presented to demonstrate the application of the proposed method.

2. THE MULTI OBJECTIVE PROBLEM

Consider a multivariate stratified population. Let there be L non-overlapping and exhaustive strata of sizes $N_1, N_2, ..., N_L$ and p characteristics be defined on each population unit. Further let the estimation of p population

means $\overline{Y}_j$; $j = 1,2,..., p$ are of interest. Unless specified otherwise in this manuscript the notations of Cochran (1977) are used. The suffix j has been introduced to denote the j^{th} characteristic.

The combined linear regression estimate of the population mean $\overline{Y}_j$ is given as

$$\bar{y}_{jlrc} = \bar{y}_{jst} + b_j\left(\overline{X}_j - \bar{x}_{jst}\right) \tag{1}$$

It can be seen that the estimate $\bar{y}_{jlrc}$ is unbiased for $\overline{Y}_j$ with a sampling variance

$$V\left(\bar{y}_{jlrc}\right) = \sum_{h=1}^{L} \frac{w_h\left(-f_h\right)}{n_h}\left(S_{yjh} - b_j S_{y_j x_j h} + b_j^2 S_{x_j h}\right) \tag{2}$$

Where b_j; j=1,2 ..., p are chosen in advance.

If the travelling costs within strata are significant, the usual linear cost function

$C = c_0 + \sum_{h=1}^{L} c_h n_h$ may be replaced by the function

$$C = c_o + \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \tag{3}$$

which is quadratic in $\sqrt{n_h}$ (See Cochran (1977) page 96), where t_h; h=1,2, ..., L are the per unit travel cost in the h^{th} stratum.

The MINLPP to be solved to obtain a compromise allocation under the discussed situation may be given as

$$\textit{Minimise} \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_p \end{pmatrix}$$

$$\textit{Subject to} \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \le C_o \tag{4}$$

$$2 \le n_h \le N_h \quad \textit{and} \quad n_h \text{ integers } ; h = 1,2..., L$$

where $V_j = V\left(\bar{y}_{jlrc}\right)$; $j = 1,2..., p$ are as given in (2) and C_o=C-c_o.

The bounded constraints $2 \le n_h \le N_h$ are introduced to check the oversampling and to have an estimate of strata standard deviations.

In the following section a procedure for solving MINLPP (4) is discussed using a Goal Programming Technique.

3. THE GOAL PROGRAMMING APPROACH

The application of goal programming technique to solve the MINLPP (4) requires the knowledge of the individual optimum allocations that are solution to the p integer nonlinear programming problems [INLPP] for each $j = 1,2...,p$

$$\begin{aligned} &\text{Minimise } V_j ; \qquad j = 1,2,...,p \\ &Subject\ to \sum_{h=1}^{L} c_h n_h + \sum_{h=1}^{L} t_h \sqrt{n_h} \le C_o \\ &2 \le n_h \le N_h \quad n_h \text{ integers; } h = 1,2,...,L \end{aligned} \tag{5}$$

Let $n_j^* = \left(\left(n_{jh}^*\right)\right)$ denote the optimal solution to (5) and V_j^* denote the optimum value of the variance V_j at $n_j^*; j = 1,2,..., p.$

Assume that the MINLPP (4) has an optimal solution

$$n_{(c)} = \left(n_{1(c),} n_{2(c),}, ..., n_{L(c)}\right) \tag{6}$$

$V_{j(c)}$ denote the value of the j^{th} variance at this solution and (c) stands for compromise.

This compromise allocation will obviously be less precise than the individual optimum allocations in the sense that $V_{j(c)} \ge V_j^*; j = 1,2..., p$.

Let $x_j \ge 0; j = 1,2,..., p$ denote the tolerance limit for the loss in precision in the j^{th} variance, that is,

$$V_{j(c)} - V_j^* \le x_j ; j = 1,2..., p \quad \text{or} \quad V_{j(c)} - x_j \le V_j^*; j = 1,2,..., p \tag{7}$$

We may now set our goal as to find a compromise allocation $n_{(c)}$ for which the total loss in precision, due to using the compromise allocation instead of the individual optimum allocations n_j^*, is minimum. The mathematical formulation of the goal programming problem [GPP] will be

$$\begin{aligned} &\textit{Minimise} \sum_{j=1}^{p} x_j \\ &\textit{Subject} \quad V_{j(c)} - x_j \le V_j^* \\ &\sum_{h=1}^{L} c_h n_{h(c)} + \sum_{h=1}^{L} t_h \sqrt{n_{h(c)}} \le C_o \\ &2 \le n_{h(c)} \le N_h \quad n_h \text{ integers } h=1,2,\ldots, L;\ j = 1,2,\ldots,p \text{ and } x_j \ge 0 \end{aligned} \tag{8}$$

The above criterion of minimizing the total loss in precision is a suitable compromise criterion to work out a compromise allocation.

4. A PRACTICAL APPLICATION

The data used for demonstrating the practical application of the developed procedure are artificially constructed with the help of the data reported in Ghufran et al. [2011]. It is assumed that the data are available for two characteristics and two auxiliary variables for a population with five strata. The total budget of the survey is taken as 1500 units out of which 300 units are used as the overhead cost

Table-1: Data for five strata, two main and two auxiliary variables

h	N_h	W_h	c_h	t_h	S_{y1h}^2	S_{y2h}^2	S_{x1h}^2	S_{x2h}^2	S_{y1x1h}	S_{y2x2h}
1	1500	0.25	1	0.5	784	42436	1444	14400	900	28900
2	1920	0.32	1	0.5	576	17689	676	33856	625	24025
3	1260	0.21	1.5	1	1024	2304	1936	29929	1089	7225
4	480	0.08	1.5	1	2916	1369	6084	8464	3481	3136
5	840	0.14	2	1.5	4489	81	5776	13689	4900	1849
Total					9789	63879	15916	100338	10995	65135

The values of b_j; $j = 1,2$ are worked out as follows [*see Cochran (1977)*].

$$b_1 = \frac{\sum_{h=1}^{5} s_{y1x1h}}{\sum_{h=1}^{5} s_{x1h}^2} == \frac{10995}{15916} = 0.69081$$

and $$b_2 = \frac{\sum_{h=1}^{5} s_{y2x2h}}{\sum_{h=1}^{5} s_{x2h}^2} = \frac{65135}{100338} = 0.64915$$

The values reported in Table 1 when substituted in INLPP [5] gives the following INLPP for *j=1*

$$\text{Minimise } V_1 = \frac{14.351175}{n_1} + \frac{3.591649}{n_2} + \frac{19.548436}{n_3} + \frac{6.463136}{n_4} + \frac{9.317236}{n_5}$$

$$\text{Subject to } n_1 + n_2 + 1.5n_3 + 1.5n_4 + 2n_5 + 0.5\sqrt{n_1} + 0.5\sqrt{n_2} + \sqrt{n_3} + \sqrt{n_4} + 1.5\sqrt{n_5} \le 1200 \quad (9)$$

$$2 \le n_1 \le 1500; 2 \le n_2 \le 1920; 2 \le n_3 \le 1260; 2 \le n_4 \le 480; 2 \le n_5 \le 840$$

$$n_h \text{ are integers}; h = 1,2,3,4,5$$

Note that the solution to the INLPP (9) will give the optimum allocation for the first characteristic. INLPP [9] is solved using the optimization software LINGO (2010). The results are:

$n_1^* = 239, n_2^* = 118, n_3^* = 224, n_4^* = 127, n_5^* = 130$ with $V_1^* = 0.3003161$.

Similarly for the second characteristics, that is, for j=2 the INLPP (5) will become

$$\text{Minimise } V_2 = \frac{686.4466}{n_1} + \frac{78.22561}{n_2} + \frac{244.1183}{n_3} + \frac{5.53070}{n_4} + \frac{67.59747}{n_5}$$

$$\text{Subject to } n_1 + n_2 + 1.5n_3 + 1.5n_4 + 2n_5 + 0.5\sqrt{n_1} + 0.5\sqrt{n_2} + \sqrt{n_3} + \sqrt{n_4} + 1.5\sqrt{n_5} \le 1200 \quad (10)$$

$$2 \le n_1 \le 1500; 2 \le n_2 \le 1920; 2 \le n_3 \le 1260; 2 \le n_4 \le 480; 2 \le n_5 \le 840$$

$$n_h \text{ are integers}; h = 1,2,3,4,5$$

The solution to INLPP (10) which is the optimum allocation for the second characteristic is

$$n_1^* = 442, n_2^* = 148, n_3^* = 212, n_4^* = 31, n_5^* = 97 \text{ with } V_2^* = 4.108390$$

Using the optimal objective values V_1^* and V_2^* from the solutions of INLPP (9) and (10) respectively, the GPP (8) will take the following form. For simplicity $n_{h(c)}$ is replaced by n_h; $h = 1,2,...,L$

$$\text{Minimise } x_1 + x_2$$

Subject to

$$\frac{14.351175}{n_1} + \frac{3.591649}{n_2} + \frac{19.548436}{n_3} + \frac{6.463136}{n_4} + \frac{9.317236}{n_5} - x_1 \le 0.300316$$

$$\frac{686.4466}{n_1} + \frac{78.22561}{n_2} + \frac{244.1183}{n_3} + \frac{5.53070}{n_4} + \frac{67.59747}{n_5} - x_2 \le 4.108390$$

$$n_1 + n_2 + 1.5n_3 + 1.5n_4 + 2n_5 + 0.5\sqrt{n_1} + 0.5\sqrt{n_2} + \sqrt{n_3} + \sqrt{n_4} + 1.5\sqrt{n_5} \le 1200$$

$$2 \le n_1 \le 1500; 2 \le n_2 \le 1920;\ 2 \le n_3 \le 1260; 2 \le n_4 \le 480;\ 2 \le n_5 \le 840$$

$$n_h \text{ are integers}; h = 1,2,3,4,5 \qquad (11)$$

The solution to the GPP (11) obtained by the optimization software LINGO [2010] is $n^*_{1(c)} = 420, n^*_{2(c)} = 144, n^*_{3(c)} = 212, n^*_{4(c)} = 45, n^*_{5(c)} = 99$

$x_1^* = 0.08874371$ and $x_2^* = 0.02644866$

5. CONCLUSION

The present section provides the summary of the results, the comparisons of the proposed allocation with some other allocations and the conclusion. For the sake of comparison we have taken in account the Cochran's Average Allocation (CAA), Cochran (1977), which is the average of the two individual allocations for j=1 and j=2.

The rounded off CAA is

$$n_1 = \frac{239 + 442}{2} = 340, n_2 = \frac{118 + 148}{2} = 133, n_3 = \frac{224 + 212}{2} = 218,$$

$$n_4 = \frac{127 + 31}{2} = 79, n_5 = \frac{130 + 97}{2} = 113.$$

Under CAA the variances for the two characteristics are found to be V_1=0.3231502 and V_2=4.3951482.

For comparing two allocations we used the Sukhatme et al. (1984) criterion. They compared different allocations with the proportional allocation in terms of the 'trace' values of the allocations. The trace value of an allocation is the sum of the variances for different characteristics under a particular allocation. The

relative efficiency [RE] of an allocation$\underline{n}$ with respect to the proportional allocation is defined as the ratio $\frac{Trace(prop)}{Trace(\underline{n})}$.

For working out the proportional allocation the total sample size n is taken as the average total sample size of the following four allocations:

i. Total sample size of the optimum allocation with respect to first characteristic = 838.

ii. Total sample size of the optimum allocation with respect to second characteristic = 930.

iii. Total sample size of the Cochran's Average Allocation = 883

iv. Total sample size of the author's proposed allocation = 920

This gives $n=\frac{838+930+883+920}{4}=\frac{3571}{4}\cong 893$

Thus the proportional allocation and the variance 'V_{prop}' under proportional allocation for the two characteristics are given as $n_h = n\,W_h = 893\,W_h$; $h=1,2,3,4,5$.

Thus n_1 = 223, n_2 = 286, n_3 = 188, n_4 = 71, n_5 = 125 and $V_{1(prop)}$= 0.346462256, $V_{2(prop)}$= 5.26893044. Table 2 and 3 give the summary of the results

Table 2: Allocations with the cost incurred

Allocations	n_h					Cost incurred
	n_1	n_2	n_3	n_4	n_5	
Proportional	223	286	188	71	125	1202.3285
Individual *j*=1	239	118	224	127	130	1200
Individual *j*=2	442	148	212	31	97	1200
CAA	340	133	218	79	113	1199.084
Proposed	420	144	212	45	99	1200

Table 3: Relative efficiency of the allocations w.r.t. proportional allocation

Allocations	Variances		Trace	Relative Efficiency (RE)
	V_1	V_2		
Proportional	0.346462256	5.26893044	5.615392696	1.00000
j=1	0.3003161	5.18843	5.4887461	1.023073866
J=2	0.453488	4.108390	4.561878	1.230938814

CAA	0.32315102	4.3951482	4.7182992	1.1901306928
Proposed	0.389059747	4.13483855	4.523898297	1.2412729746

It can be seen that the author's proposed allocation is the most efficient among the considered allocations in terms of the relative efficiency (RE). It is also to be noted that the compromise allocation is nearly as good as the individual optimum allocation for the second characteristic.

ACKNOWLEDGEMENT

The author Mohammad Jameel Ahsan is grateful to the UGC for its financial support in the form of Emeritus Fellowship in the preparation of this manuscript.

REFERENCES

Ansari, A. H., Najmussehar and Ahsan, M. J. (2009): On multiple response stratified random sampling design, *journal of statistics sciences*, 1(1), 45-54.

Ansari, A. H., Varshney,R., Najmussehar and Ahsan, M.J. (2011): An Optimum multivariate – multiobjective stratified sampling design, *Metron, LXIX* (3), 227-250.

Ahsan, M. J. and Khan, S. U. (1977): Optimum allocation in multivariate stratified random sampling using prior information, *Journal of Indian Statistical Association*, 15, 57-67.

Ahsan, M.J. and Khan, S. U. (1982): Optimum allocation in multivariate stratified random sampling with overhead cost, *Metrika*, 29, 71-78.

Bethal, J. (1985): An Optimum allocation algorithm for multivariate surveys, Proceedings of the Survey Research Methods section, *American Statistical Association*, 209-212.

Bethal, J. (1989): Sample allocation in multivariate surveys, *Survey Methodology*, 15, 47-57.

Boer, E. P.J. and Hendrix, E. M. T. (2000): Global Optimization problem in optimal design of experiment in regression models, *J. Global Optim.* , 18(4), 385-398.

Chatterjee, S. (1967): A note on optimum allocation, *Scandinavian Actuarial Journal*, 50, 40-44.

Chromy, J. R. (1987): Design optimization with multiple objectives, Proceedings of the Survey Research Methods section, *American Statistical Association*, 194-199.

Chinchuluun, A. and Pardalos, P. M. (2007): A survey of recent developments in multiobjective optimization, *Ann. Operational Research*, 154(1), 29-50.

Cochran, W. G. (1977): *Sampling Techniques*, third edition, John Wiley and Sons, NewYork.

Dayal, S. (1985): Allocation of sample using values of Auxiliary characteristic, *Journal of statistical planning and inference*, 11, 321-328.

Dalenius, T. (1957): Sampling in Sweden, *Contributions to the Methods and Theories of Sample Survey Practice,* Almqvist and Wicksell, Stockholm.

Diaz Garcia, J. A. and Garay Tapia, M. M. (2007): "Optimum allocation in Stratified Surveys: Stochastic Programming", *Computational Statistics and Data Analysis*, Vol. 51, No. 6, pp. 3016-3026.

Floudas, C. A. and Pardalos, P. M. (2009): *Encyclopedia of Optimization*, Springer, Berlin.

Ghufran, S., Khowaja, S. and Ahsan, M. J. (2011): Multiobjective optimum allocation problem with probabilistic nonlinear cost constraint, *International Journal of Engineering Science and Technology*, Vol.3, No. 6, 135-145.

Hendrix, E. M. T., Ortigosa, P. M. and Garcia, I. (2001): On Success rates for controlled random search, *J. Global Optim*., 21 (3), 239-263.

Holmberg, A. (2003): A multiparameter perspective on the choice of sampling design in surveys, *Statistics in Transition*, 5 (6), 969-994.

Ige, A. F. and Tripathi, T. P. (1987): On double sampling for stratification and use of auxiliary information, *J. Indian Soc. Agricultural statistics*, 39 (2), 191-201.

Jahan, N., Khan, M. G. M. and Ahsan, M. J. (1994): A generalized compromise allocation, *Journal of the Indian Statistical Association*, 32, 95-101.

Khan, M. G. M., Ahsan, M.J. and Jahan, N. (1997): Compromise allocation in multivariate stratified sampling: an integer solution, *Naval Research Logistics*, 44, 69-79.

Khan, M. G. M., Khan, E. A. and Ahsan, M. J. (2003): An Optimal multivariate stratified sampling design using dynamic programming, *Australian and NewZealand J. Statist*. , 45 (1), 107-113.

Khan, M. G. M., Maiti, T. and Ahsan, M. J. (2010): An Optimal multivariate stratified sampling design using auxiliary information: an integer solution using goal programming approach, *Journal of official statistics*, 26 (4), 695-708.

Kozak, M. (2004): Optimal stratification using random search method in agricultural surveys, *Statistics in Transition*, 6 (5), 797-806.

Kozak, M. (2006): On Sample allocation in multivariate surveys, *communication in Statistics- Simulation and Computation*, 35, 901-910.

Kokan, A. R. and Khan, S.U. (1967): Optimum allocation in multivariate surveys: An analytical solution, *Journal of Royal Statistical Society*, Ser. B, 29, 115-125.

Lingo User's Guide (2010), published by Lindo Systems Inc., 1415 North Dayton Street, Chicago, Illinois-60622 (USA).

Neyman, J. (1934): On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection, *Journal of Royal Statistical Society*, 97, 558-625.

Najmussehar, Ahsan and Khan (2005): Allocation of a sample to strata: The multivariate case, *Pure and Applied Mathematika Sciences*, Vol. LXII, No. 1-2, 1-22.

Najmussehar and Bari, A. (2002): Double sampling for stratification with sub sampling the non respondents: a dynamic programming approach. *Aligarh J. Statist*., 22, 27-41.

Pardalos, P. M., Siskos, Y. and Zopounidis, C. (eds.) (1995): Advances in multicriteria analysis, *Kluwer Academic Publishers,* Dordrecht.

Rao, J. N. K. (1973): On double sampling for stratification and analytical surveys. *Biometrika*, 60, 125-133.

Tripathi, T. P. and Bahl, S. (1991): Estimation of mean using double sampling for stratification and multivariate auxiliary information. *Comm. Statist. Theory Methods*, 20 (8), 2589-2602.

Yates, F. (1960): *Sampling Methods for censuses and Surveys*, 3rd ed., Charles Griffin and Co. Ltd., London.

Zopounidis, C. and Pardalos, P. M. (eds.) (2010): *Handbook of Multicriteria Analysis*, Springer, Berlin.

Farha Naz, Ummatul Fatima,
M. J. Ahsan and Q. M. Ali
Department of Statistics & O.R.,
Aligarh Muslim University,
Aligarh-202002, India.
E-mail: *farhajeelani16@gmail.com*

Received: 31.03.2015

Revised: 07.05.2015