STATISTICS IN TRANSITION-new series, Spring 2013
Vol. 14, No. 1, pp. 31–44

SYNTHETIC ESTIMATORS USING AUXILIARY INFORMATION IN SMALL DOMAINS

P. K. Rai[1], K. K. Pandey[2]

ABSTRACT

In the present article we discuss the generalized class of synthetic estimators for estimating the population mean of small domains under the information of two auxiliary variables, and describe the special cases under the different values of the constant beta involved in the proposed generalized class of synthetic estimator. In addition we have taken a numerical illustration for the two auxiliary variables and compared the result for the synthetic ratio estimator under single and two auxiliary variables.

Key words: auxiliary information, small area (domain) estimation, synthetic estimation, optimum weights.

1. Introduction

An estimator is called a synthetic estimator if a reliable direct estimator for a larger area, covering several small areas, is used to derive an indirect estimator for a small area under the assumption that the small areas have the same characteristics as the large area (Gonzalez, 1973). Such estimators have been studied by Gonzalez (1973), Gonzalez and Waksberg (1973). It is a fact that if small domain sample sizes are relatively small the synthetic estimator performs better than the simple direct estimators, whereas when sample sizes are large the direct estimators perform better than the synthetic estimators (Schaible, Brock, Casady and Schnack, 1977). The classes of synthetic estimators proposed by the above authors give consistent estimators if the corresponding synthetic assumptions are satisfied. These authors, further, discuss the generalized class of synthetic estimators under simple random sampling and stratified random

[1] Department of Mathematics and Statistics, Banasthali University, Rajasthan India-304022. E-mail: raipiyush5@gmail.com.
[2] College of management & Economic Studies, University of Petroleum & Energy Studies (UPES), Energy Acres, P.O. Bidholi, Dehradun-248007. E-mail: krishan.pandey@gmail.com, kkpandey@ddn.upes.ac.in.

sampling schemes. In sample surveys usually auxiliary variables are used to increase the precision of the estimators. A ratio estimator is one of the most commonly used estimators among others for the population mean or population total with the help of an auxiliary character. It was shown by Tikkiwal, G.C. and Ghiya, A. (2004), Tikkiwal, G.C. and Pandey, K.K. (2007), Pandey Krishan K. and Tikkiwal, G.C. (2010), Pandey, Krishan K. (2010), that when an auxiliary variable is closely related with the variable under study, the small area estimators based on auxiliary information perform better than those which do not use auxiliary information. Further, Tikkiwal, G.C. and Pandey, K.K. (2007) discuss the generalized class of synthetic and composite estimators under Lahiri-Midzuno and systematic sampling schemes. The relative performances of these estimators are empirically assessed for the problem of crop acreage estimation for small domains.

It is rather difficult to assess the performance of these estimators theoretically. Here we have discussed the different aspect of the generalized class of synthetic estimators for small area estimation problems when more than one auxiliary information is available.

2. Generalized class of synthetic estimators in sample surveys

We define a generalized class of synthetic estimators for estimating the population mean $\overline{Y}$ under ‘k’ auxiliary variables $x_1, x_2, \ldots\ldots. x_k$, as follows

$$\overline{y}_{syn} = \sum_{i=1}^{k} W_i \overline{y} \left(\frac{\overline{x}_i}{\overline{X}_i} \right)^{\beta_i} \tag{1}$$

where β_i's are equal to the $-\rho_{0i} \frac{C_0}{C_i}$ and W_i are the weights to be obtained by minimizing the variance of (1) subject to the condition that $\sum_{i=1}^{k} W_i = 1$. Here, $\overline{x}_i$ and $\overline{X}_i$ denote the sample mean and population mean of $x_i (i = 1, 2, \ldots.., p)$ respectively, $\rho_{ij} (i \neq j = 0, 1, \ldots.., p)$ denotes the correlation coefficient between x_i and x_j, and $C_i (i = 0, 1, \ldots.., p)$ denotes the coefficient of variation of x_i; the suffix 0 stands for the variable y and $\overline{y}$ is the sample mean of the variable under study.

3. Notations and formulation under small domains

Let us represent the important notations which are to be used in this paper. Suppose that a finite population U= (1,…, i,…, N) is divided into 'A' non-overlapping domains U_a of size N_a (a=1,…, A) for which estimates are required. The domains may be numerous and represent small geographical areas of a sampled population, which may be a state or a sub-division of the state as the case may be. Let the characteristic under study be denoted by 'y'. Further, assume that the auxiliary information is also available and denoted by 'x'. A simple random sample (without replacement) s= (1,…, i, …, n) of size n is selected such that n_a (a=1,…, A) units in the sample 's' come from small area 'a'. Consequently,

$$\sum_{a=1}^{A} N_a = N \text{ and } \sum_{a=1}^{A} n_a = n \tag{2}$$

Let us consider the case of generalized synthetic estimator for estimating the population mean $\overline{Y}_a$ for domain 'a' under two auxiliary variables x_1 and x_2;

$$\overline{y}_{syn,a} = W_1\overline{y}\left(\frac{\overline{x}_1}{\overline{X}_{1a}}\right)^{\beta_1} + W_2\overline{y}\left(\frac{\overline{x}_2}{\overline{X}_{2a}}\right)^{\beta_2} \tag{3}$$

Here, W_1 and W_2 are the weights such that $W_1 + W_2 = 1$ and β_1 , β_1 are suitably chosen constants. To find the expectation and mean square error of the estimator $\overline{y}_{syn,a}$ define

$$\varepsilon_0 = \frac{\overline{y} - \overline{Y}}{\overline{Y}}, \qquad \varepsilon_1 = \frac{\overline{x}_1 - \overline{X}_1}{\overline{X}_1}, \qquad \varepsilon_2 = \frac{\overline{x}_2 - \overline{X}_2}{\overline{X}_2} \tag{4}$$

Then, clearly

$$E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_2) = 0 \tag{5}$$

$$E(\varepsilon_0^2) = \frac{f}{n}C_0^2, \quad E(\varepsilon_1^2) = \frac{f}{n}C_1^2, \quad E(\varepsilon_2^2) = \frac{f}{n}C_2^2 \tag{6}$$

and $$E(\varepsilon_o\varepsilon_1) = \frac{f}{n}C_{01}, \qquad E(\varepsilon_o\varepsilon_2) = \frac{f}{n}C_{02}, \qquad E(\varepsilon_1\varepsilon_2) = \frac{f}{n}C_{12} \tag{7}$$

where

$$f = \frac{N-n}{N}, \qquad C_0^2 = \frac{S_y^2}{\overline{Y}^2}, \qquad C_1^2 = \frac{S_{x_1}^2}{\overline{X}_1^2}, \qquad C_2^2 = \frac{S_{x_2}^2}{\overline{X}_1^2},$$

$$C_{01} = \frac{S_{yx_1}}{\overline{Y}\,\overline{X}_1}, \quad C_{02} = \frac{S_{yx_2}}{\overline{Y}\,\overline{X}_2}, \quad C_{12} = \frac{S_{x_1x_2}}{\overline{X}_1\overline{X}_2}, \tag{8}$$

and

$$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{Y})^2 ; \qquad S_{x_1}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_{1i} - \overline{X}_1)^2$$

$$S_{x_2}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_{2i} - \overline{X}_2)^2 ;$$

$$S_{yx_1} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{Y})(x_{1i} - \overline{X}_1)$$

$$S_{yx_2} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{Y})(x_{2i} - \overline{X}_2)$$

$$S_{x_1x_2} = \frac{1}{N-1}\sum_{i=1}^{N}(x_{1i} - \overline{X}_1)(x_{2i} - \overline{X}_2) \tag{9}$$

4. Bias and mean square error

In this section bias and mean square error expressions are considered up to the terms of order (1/n) only. The $\overline{y}_{syn,a}$ can be expressed as

$$\overline{y}_{syn,a} = W_1\overline{Y}(1+\varepsilon_0)\left(\frac{\overline{X}_1(1+\varepsilon_1)}{\overline{X}_{1a}}\right)^{\beta_1} + W_2\overline{Y}(1+\varepsilon_0)\left(\frac{\overline{X}_2(1+\varepsilon_2)}{\overline{X}_{2a}}\right)^{\beta_2} \tag{10}$$

assuming that the contribution of terms involving powers in $\varepsilon_0, \varepsilon_1$ and ε_2 higher than the second order is negligible. The design bias of $\bar{y}_{syn,a}$ and $MSE(\bar{y}_{syn,a})$ is given below as

$$B(\bar{y}_{syn,a}) = W_1\bar{Y}\left(\frac{\bar{X}_1}{\bar{X}_{1a}}\right)^{\beta_1}\left(1+\frac{f}{n}\left(\frac{\beta_1(\beta_1-1)}{2!}C_1^2+\beta_1 C_{01}\right)\right) + W_2\bar{Y}\left(\frac{\bar{X}_2}{\bar{X}_{2a}}\right)^{\beta_2}\left(1+\frac{f}{n}\left(\frac{\beta_2(\beta_2-1)}{2!}C_2^2+\beta_2 C_{02}\right)\right) - \bar{Y}_a \tag{11}$$

$$\begin{aligned} MSE(\bar{y}_{syn,a}) &= \bar{Y}^2\left[W_1\left(\frac{\bar{X}_1}{\bar{X}_{1a}}\right)^{\beta_1} + W_2\left(\frac{\bar{X}_2}{\bar{X}_{2a}}\right)^{\beta_2}\right]^2 \\ &+ \bar{Y}^2W_1^2\left(\frac{\bar{X}_1}{\bar{X}_{1a}}\right)^{2\beta_1}\left[\frac{f}{n}\left\{C_0^2+\beta_1(2\beta_1C_1^2-C_1^2+4C_{01})\right\}\right] \\ &+ \bar{Y}^2W_2^2\left(\frac{\bar{X}_2}{\bar{X}_{2a}}\right)^{2\beta_2}\left[\frac{f}{n}\left\{C_0^2+\beta_2(2\beta_2C_2^2-C_2^2+4C_{02})\right\}\right] \\ &+ 2W_1W_2\bar{Y}^2\left(\frac{\bar{X}_1}{\bar{X}_{1a}}\right)^{\beta_1}\left(\frac{\bar{X}_2}{\bar{X}_{2a}}\right)^{\beta_2}\left(\frac{f}{n}\left\{\begin{array}{l} C_0^2+\beta_1\left(2C_{01}+\frac{(\beta_1-1)}{2!}C_1^2\right) \\ +\beta_2\left(2C_{02}+\frac{(\beta_2-1)}{2!}C_2^2\right)+\beta_1\beta_2C_{12}\end{array}\right\}\right) \\ &- 2\bar{Y}_a\left[\begin{array}{l} W_1\bar{Y}\left(\frac{\bar{X}_1}{\bar{X}_{1a}}\right)^{\beta_1}\left(1+\frac{f}{n}\left(\frac{\beta_1(\beta_1-1)}{2!}C_1^2+\beta_1C_{01}\right)\right) \\ +W_2\bar{Y}\left(\frac{\bar{X}_2}{\bar{X}_{2a}}\right)^{\beta_2}\left(1+\frac{f}{n}\left(\frac{\beta_2(\beta_2-1)}{2!}C_2^2+\beta_2C_{02}\right)\right)\end{array}\right] + \bar{Y}_a^2 \end{aligned} \tag{12}$$

Also, the optimum value for the weights W_1^{opt} and W_2^{opt} can be obtained by minimizing mean square error term of (12).

5. Special cases: various synthetic estimators

The generalized synthetic estimator $\bar{y}_{syn,a}$ reduces to the simple synthetic estimator if β_1 and β_2 equal to zero, i.e. $\beta_1 = \beta_2 = 0$

$$\bar{y}_{syn,a} = \bar{y} = \bar{y}_{syn,s,a} \tag{13}$$

and synthetic assumption $\bar{Y}_a\left(\bar{X}_a\right)^{\beta} \cong \bar{Y}(\bar{X})^{\beta}$ reduces to $\bar{Y}_a \cong \bar{Y}$. Substituting $\beta_1 = \beta_2 = 0$ in the expression (11) we get

$$B(\bar{y}_{syn,a}) = W_1\bar{Y} + W_2\bar{Y} - \bar{Y}_a = \bar{Y} - \bar{Y}_a = B(\bar{y}_{syn,s,a}) \tag{14}$$

This is the expression for design bias of the simple synthetic estimator. The design bias of the synthetic estimator vanishes if the synthetic assumption, i.e. $\bar{Y}_a \cong \bar{Y}$ is satisfied. Now $\beta_1 = \beta_2 = 0$ in the expression (12) gives

$$MSE(\bar{y}_{syn,s,a}) = \bar{Y}^2\left(W_1 + W_2\right)^2 + \bar{Y}^2\frac{f}{n}C_0^2\left(W_1 + W_2\right)^2 - 2\bar{Y}_a\bar{Y}(W_1 + W_2) + \bar{Y}_a^2$$

$$= \bar{Y}^2\frac{f}{n}C_0^2 = \frac{N-n}{Nn}S_y^2 \tag{15}$$

This is the mean square error of simple synthetic estimator under said synthetic assumption.

The generalized Synthetic estimator $\bar{y}_{syn,a}$ reduces to ratio synthetic estimator under two auxiliary variables, if β_1 and β_2 equal to -1, i.e. $\beta_1 = \beta_2 = -1$

$$\bar{y}_{syn,r,a} = W_1\left(\frac{\bar{y}}{\bar{x}_1}\right)\bar{X}_{1a} + W_2\left(\frac{\bar{y}}{\bar{x}_2}\right)\bar{X}_{2a} \tag{16}$$

Substituting $\beta_1 = \beta_2 = -1$ in the expression (11) and (12) we get the expressions for the bias and mse for the ratio synthetic estimator.

The generalized synthetic estimator $\bar{y}_{syn,a}$ reduces to the product synthetic estimator under two auxiliary variables, if β_1 and β_2 equal to +1, i.e. $\beta_1 = \beta_2 = +1$

$$\bar{y}_{syn,p,a} = W_1\bar{y}\left(\frac{\bar{x}_1}{\bar{X}_{1a}}\right) + W_2\bar{y}\left(\frac{\bar{x}_2}{\bar{X}_{2a}}\right) \tag{17}$$

Substituting $\beta_1 = \beta_2 = +1$ in the expression (11) and (12) we get the expressions for the bias and mse for the product synthetic estimator.

6. Numerical illustration

We consider the study variable as REV84, the real estate values according to 1984 assessment and use the two auxiliary variables as population under municipalities of 1975 and 1985 of the different geographic region indicator of Swedish municipalities. Just draw the sample of different sizes using SRSWOR scheme and analyze the cases for regions 1, 2 and 3 as small domains.

We have considered the cases under single and double auxiliary variables and computed the biases and mse's for the different sample sizes. Using the expressions of optimum weights we have computed the value of weights for the generalized synthetic estimator under $\beta_1 = \beta_2 = -1$ which reduces to synthetic ratio estimator, thus W_1^{opt} = 0.978828466 and W_2^{opt} =0.021171534. And $\overline{Y}_a$= 3011.683, $\overline{X}_{1a}$=28.92308, $\overline{X}_{2a}$= 255.0192, $\overline{Y}$=3133.862676, $\overline{X}_1$= 28.80986, $\overline{X}_2$= 111.9471831

Under single auxiliary variable the bias and mse for the synthetic ratio estimator is given by

$$B_2 = B\left(\overline{y}_{syn,r,a}\right) = \frac{\overline{Y}}{\overline{X}}\overline{X}_a\left[1+\frac{N-n}{Nn}\left(C_x^2 - C_{xy}\right)\right] - \overline{Y}_a \tag{18}$$

and
$$MSE\left(\overline{y}_{syn,r,a}\right) = \left(\frac{\overline{Y}}{\overline{X}}\overline{X}_a\right)^2\left[1+\frac{N-n}{Nn}\left\{3C_x^2 + C_y^2 - 4C_{xy}\right\}\right] -2\overline{Y}_a\left(\frac{\overline{Y}}{\overline{X}}\overline{X}_a\right)\left[1+\frac{N-n}{Nn}\left(C_x^2 - C_{xy}\right)\right] + \overline{Y}_a^2 \tag{19}$$

Using the equations above we show the results for bias and mse of synthetic ratio estimator under two scenarios for the different sample sizes in the given tables in appendix. The results can be also explored by the following graphical presentation.

7. Conclusions

At least two auxiliary variables will be the better choice over a single one when the sample size decreases. In sample surveys it is useful to make use of information on the auxiliary variable to increase the precision of the estimators. The above study will provide the motivation towards the use of generalized class of synthetic estimators in the small area estimation, when the information on two auxiliary variables is available.

Acknowledgements

The authors are grateful to the University Grant Commission (UGC), F.No.36-341/2008 (SR), New Delhi for providing support and facilitation to this research and development work.

REFERENCES

AGRAWAL, M. C. and ROY, D. C., (1997). Efficient Estimators for Small Domains. Jour. Ind. Soc. Ag. Statistics 52(3), 327-337.

GHOSH, M. and RAO, J. N. K., (1994). Small Area Estimation: An Appraisal. Statistical Science, 9, 55-93.

GONZALEZ, M. E., (1973). Use and Evaluation of Synthetic Estimates. Proceedings of the Social Statistical Section of American Statistical Association, 33-36.

GONZALEZ, M. E. and WAKSBERG, J., (1973). Estimation of the Error of Synthetic Estimates. Paper presented at first meeting of the International Association of Survey Statisticians, Vienna, Austria, 18-25, August 1973.

HEDAYAT, A. S. and SINHA, B. K., (1991). Design and Inference in Finite Population Sampling. John Wiley and Sons, New York.

PANDEY, KRISHAN K. and TIKKIWAL, G. C., (2010). „Generalized class of synthetic estimators for small area under systematic sampling design" Statistics in Transition-new series, Poland, Vol.11 No.1, pp. 75-89.

PANDEY, KRISHAN, K., (2010). „Aspects of small area estimation using auxiliary information". Book published by VDM Verlag Dr. Muller GmbH & Co. KG, Germany. (ISBN: 978-3-639-31569-1).

PLATEK, R., RAO, J. N. K., SARNDAL, C. E. and SINGH, M. P., (1987). Small Area Statistics: An International Symposium. John Wiley and Sons, New York.

PURCELL, N. J. and KISH, L., (1979). Estimation for Small Domain. Biometrics, 35, 365-384.

SARNDAL, C. E., SWENSSON, B. and WRETMAN, J., (1992). Model Assisted Survey Sampling. Springer-Verlag, New York.

SCHAIBLE, W. L., BROCK, D. B., CASADY, R. J. and SCHNACK, G. A., (1977). An Empirical Comparison of the Simple Inflation, Synthetic and Composite Estimators for Small Area Statistics. Proceedings of the American Statistical Association, Social Statistics Section, 1017-1021.

SINGH, M. P., GAMBINO, J. and MANTEL, H., (1993). Issues and Options in the Provision of Small Area Data. Proceedings of International Scientific Conference on Small Area Statistics and survey Design (Held in September, 1992 in Warsaw, Poland), 37-75.

SRIVASTAVA, S. K., (1967). An estimator using auxiliary information in sample surveys, Calcutta Statistical Association Bulletin, 16, 121-132.

TIKKIWAL, B. D. and TIKKIWAL, G. C., (1991). Sampling Strategies in Surveys. The Role of the Theory of T-Classes and Computers. Symposium Ind. Agri. Stat. Research Institute, New Delhi, 287-296.

TIKKIWAL, G. C. and GUPTA, A. K., (1991). Estimation of Population Mean under Successive Sampling When Various Weights and Regression Coefficient are Unknown. Biometrical Journal, 33, 529-538.

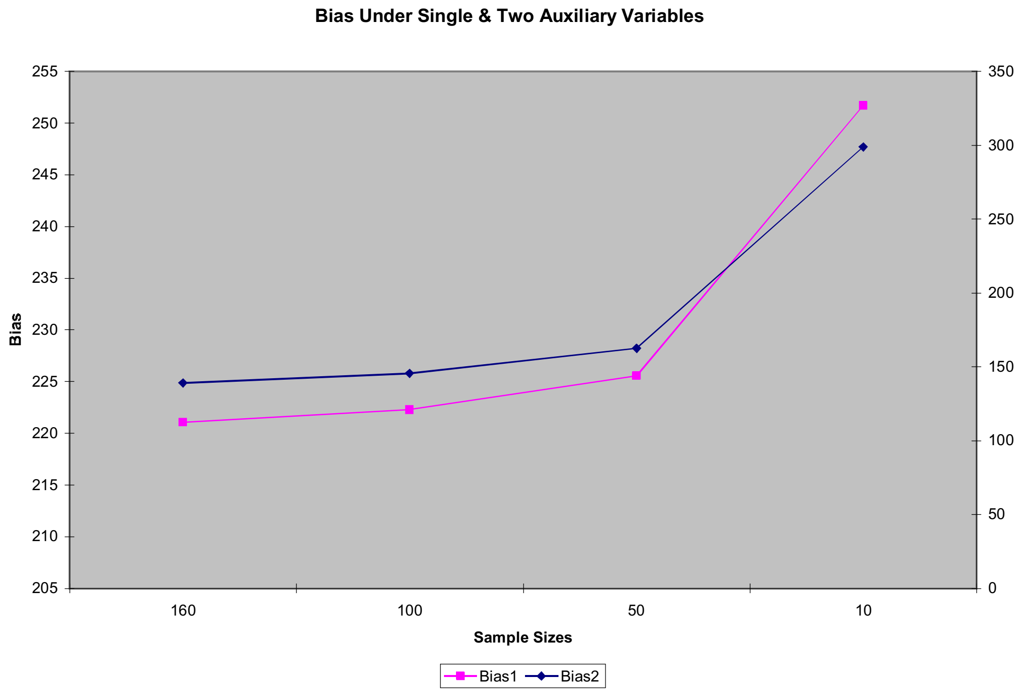
TIKKIWAL, G. C. and GHIYA, ALKA., (2000). A Generalized Class of Synthetic Estimators with Application to Crop Acreage Estimation for Small Domains. Biometrical Journal, 42 (7), 865-876.

TIKKIWAL, G. C. and PANDEY, K. K., (2007). On Some Aspects of Small Area Estimation Using Auxiliary Information. Ph.D. Thesis under supervision of Prof. G. C. Tikkiwal, Head of Department of Mathematics and Statistics J. N. V. University Jodhpur Rajasthan.

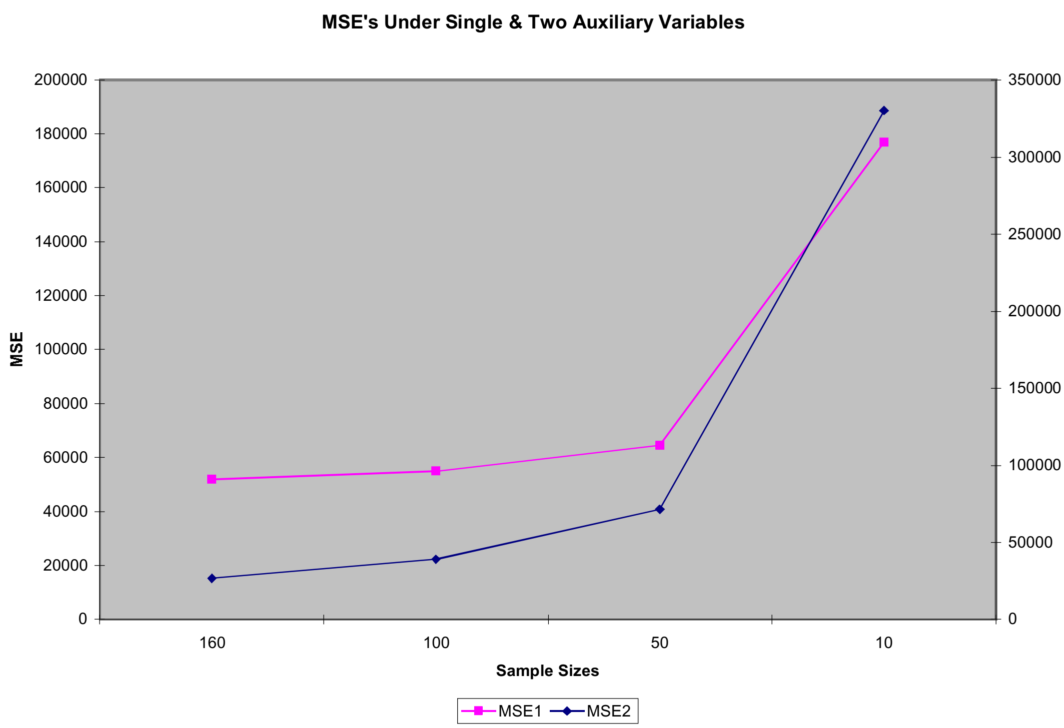
APPENDICES

FIGURES

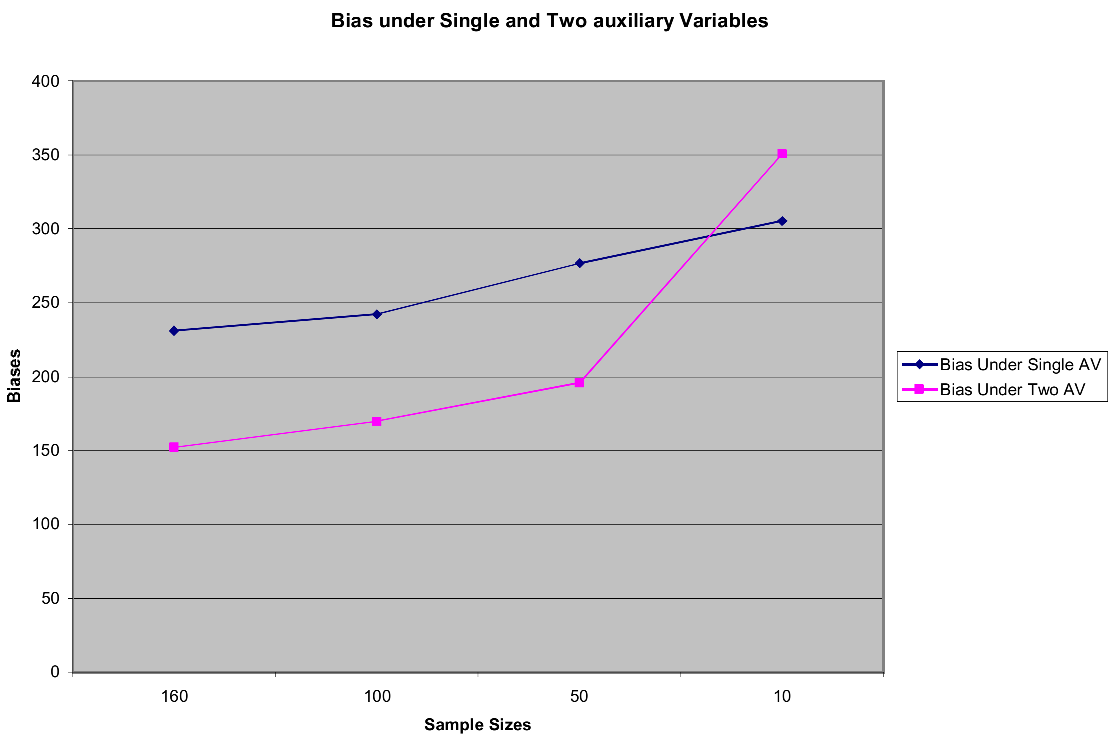
Graph 1.



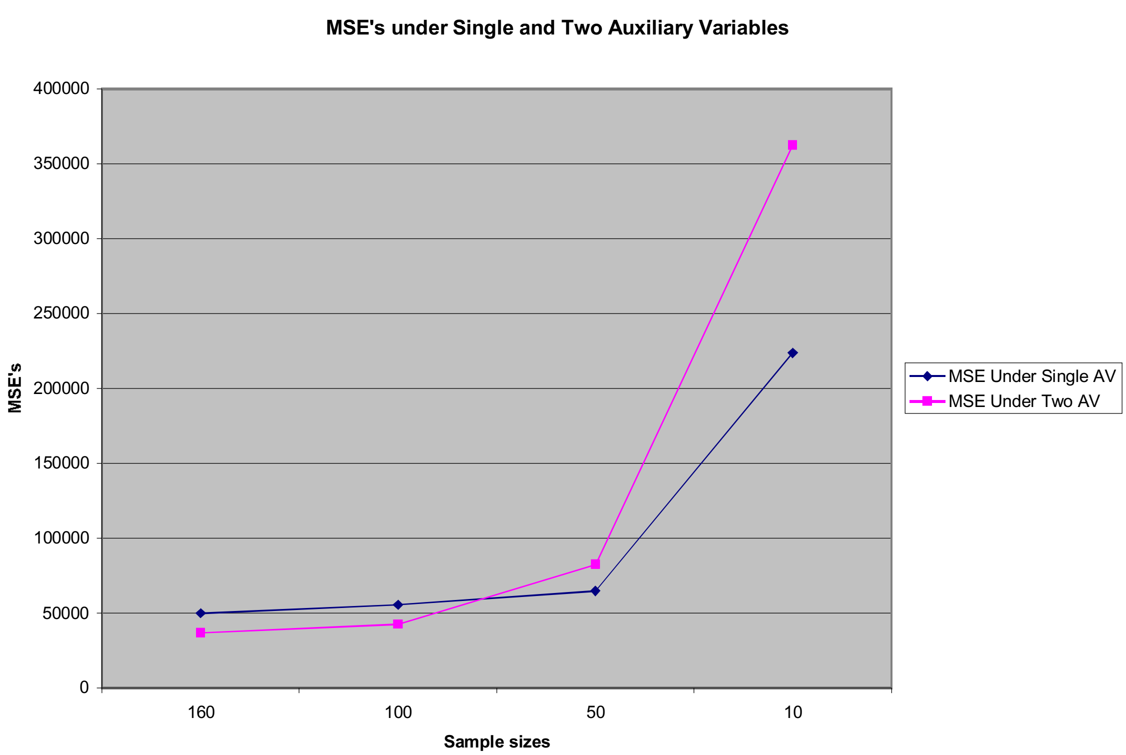
Graph 2.



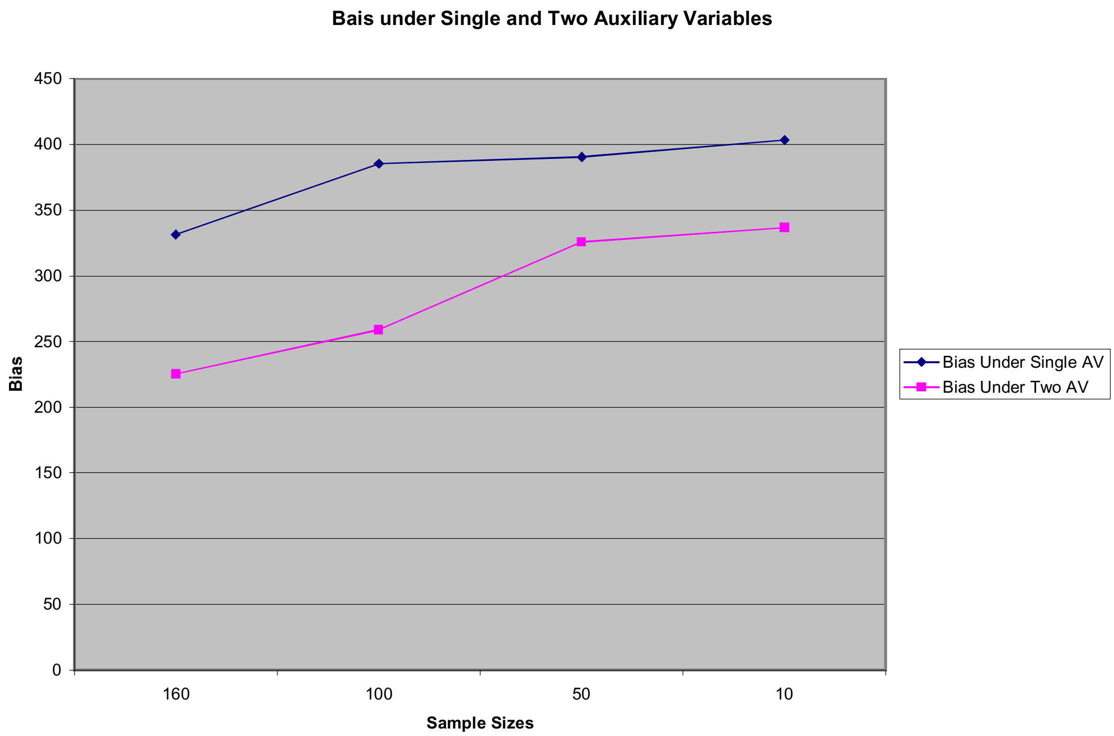
Graph 3.



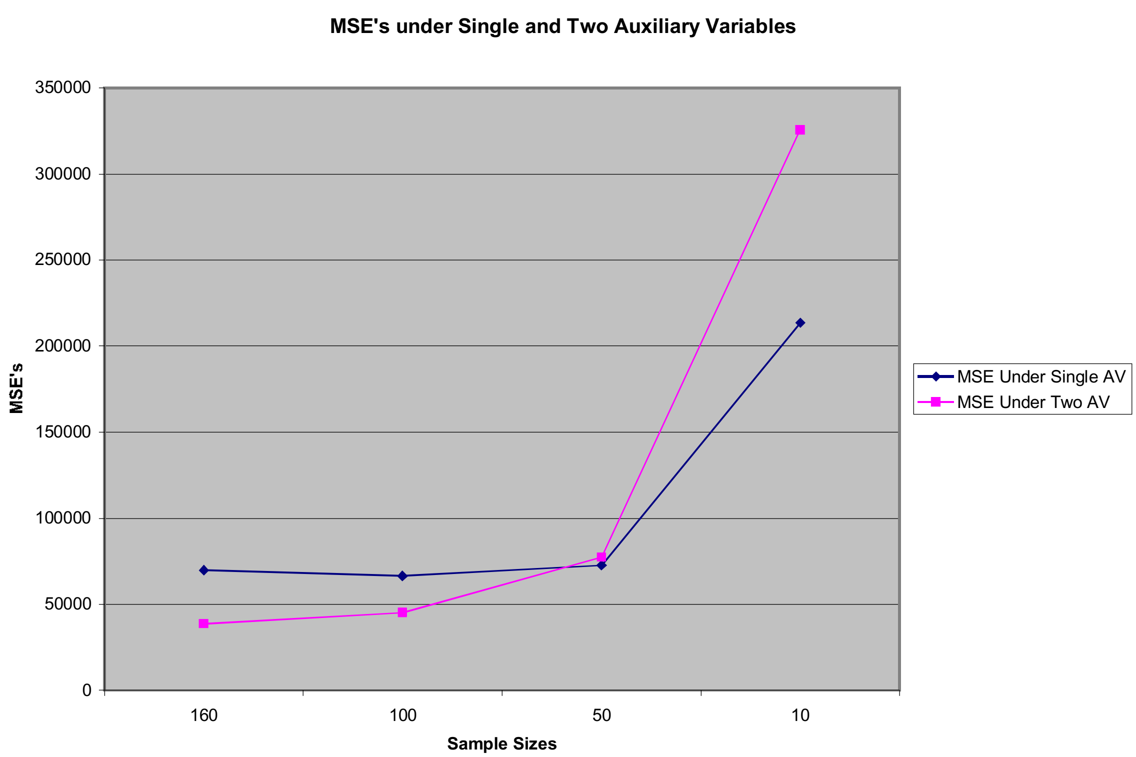
Graph 4.



Graph 5.



Graph 6.



TABLES

Table No. 6.1
Coefficient of Variation involved in Computation

C_0^2	**2.54**
C_1^2	**3.34**
C_2^2	**2.81**
C_{01}	**2.81**
C_{02}	**1.56**
C_{12}	**1.87**

Table No. 6.2
Bias of Synthetic Ratio Estimator under Single and Two Auxiliary Variables

Sample Sizes	Bias Under Single AV	Bias Under Two AV
160	221.07	139.14
100	222.30	145.53
50	225.57	162.56
10	251.71	298.79

Table No. 6.3
MSE of Synthetic Ratio Estimator under Single and Two Auxiliary Variables

Sample Sizes	MSE Under Single AV	MSE Under Two AV
160	51908.99	26914.89
100	54960.31	39043.23
50	64560.66	71385.74
10	176930.48	330125.22

Table No. 6.4
Bias of Synthetic Ratio Estimator under Single and Two Auxiliary Variables

Sample Sizes	Bias Under Single AV	Bias Under Two AV
160	231.13	152.10
100	242.29	169.76
50	276.56	195.83
10	305.31	350.77

Table No. 6.5
MSE of Synthetic Ratio Estimator under Single and Two Auxiliary Variables

Sample Sizes	MSE Under Single AV	MSE Under Two AV
160	50025.81	36980.87
100	55635.45	42565.40
50	64860.68	82344.91
10	223692.88	362512.46

Table No. 6.6
Bias of Synthetic Ratio Estimator under Single and Two Auxiliary Variables

Sample Sizes	Bias Under Single AV	Bias Under Two AV
160	331.74	225.37
100	385.35	258.88
50	390.56	325.83
10	403.30	336.73

Table No. 6.7
MSE of Synthetic Ratio Estimator under Single and Two Auxiliary Variables

Sample Sizes	MSE Under Single AV	MSE Under Two AV
160	69871.32	38565.61
100	66523.33	45021.33
50	72563.24	77452.26
10	213564.13	325613.65