World Applied Sciences Journal 7 (5): 593-596, 2009
ISSN 1818-4952

A General Shrinkage Estimator in Survey Sampling

[1]Muhammad Qaiser Shahbaz and [2]Muhammad Hanif

[1]Department of Mathematics, COMSATS Institute of Information Technology, Lahore, Pakistan
[2]Department of Mathematics, LUMS, Lahore, Pakistan

Abstract: The shrinkage estimators are widely used in estimation processes. These estimators have smaller Mean Square Error's or Variance as compared with the conventional estimators. In this article we have proposed a general class of shrinkage estimators for estimation of any population characteristic. The Mean Square Error of the proposed estimator has also been developed. Some examples are given for practical applicability of the proposed estimator.

Key words: Mean square error . shrinkage estimator . population characteristics . regression estimator

INTRODUCTION

Survey statisticians have always been interested in estimation of certain population characteristics based upon the sampled information. The core objective in estimation is to obtain the precise result with respect to smaller Mean Square Error (MSE) of the estimator used. The estimators proposed are generally based upon the sample of size *n* drawn from a population of size *N* with units identified as U_1, U_2,…,U_N. The population characteristics like Mean, Proportions, total are estimated on the basis of sample data.

The most classical estimator in the context of survey sampling is the estimator of population mean. This estimator is given as:

$$\bar{y} = n^{-1}\sum_{i=1}^{n} y_i \tag{1.1}$$

The variance of (1.1) is given as:

$$\mathrm{Var}(\bar{y}) = \frac{N-n}{Nn}S_y^2 = \theta S_y^2 \tag{1.2}$$

with $\theta = n^{-1} - N^{-1}$. The use of auxiliary information has always been useful in increasing the precision of estimates. Utilizing the information on one auxiliary variable *X*, Hansen *et al*. [2] has proposed the following regression estimator for estimation of population mean:

$$\bar{y}_{lr} = \bar{y} + \beta\left(\bar{X} - \bar{x}\right) \tag{1.3}$$

The estimator (1.3) has the smaller variance as compared with (1.1). The variance of (1.3) is given as:

$$\mathrm{Var}\left(\bar{y}_{lr}\right) = \frac{N-n}{Nn}S_y^2\left(1-\rho^2\right) = \theta S_y^2\left(1-\rho^2\right) \tag{1.4}$$

where ρ is the correlation coefficient between *X* and *Y*. Clearly, the variance of (1.3) is smaller than the variance of (1.1). The reduction in variance depends upon the strength of the correlation coefficient between *X* and *Y*. The estimator (1.3) has been extensively used in single and two phase sampling. The regression estimator for two phase sampling, given by Cochran [1], has the form:

$$\bar{y}_{1(2)} = \bar{y}_2 + \beta\left(\bar{x}_1 - \bar{x}_2\right) \tag{1.5}$$

where $\bar{y}_2$ is mean of variable of interest for second phase sample. The variance of (1.5) is given as:

$$\mathrm{Var}\left(\bar{y}_{1(2)}\right) = S_y^2\left\{\theta_2\left(1-\rho^2\right) + \theta_1\rho^2\right\} \tag{1.6}$$

where $\theta_j = n_j^{-1} - N^{-1}$ and n_j is the sample size at *j*–th phase. Multiple auxiliary variables have also been used in estimation process. Kiregyera [3] has proposed the following regression–in–regression type estimator for two phase sampling using two auxiliary variables:

$$\bar{y}_{K(2)} = \bar{y}_2 + \beta_{yx}\left\{\left(\bar{x}_1 - \bar{x}_2\right) + \beta_{xz}\left(\bar{Z} - \bar{z}_1\right)\right\} \tag{1.7}$$

The variance of (1.7) is given as:

$$\mathrm{Var}\left(\bar{y}_{K(2)}\right) = S_y^2\left\{\begin{array}{l}\theta_2 - \left(\theta_2 - \theta_1\right)\rho_{xy}^2 - \\ \theta_1\rho_{yz}^2 + \theta_1\left(\rho_{yz} - \rho_{xy}\rho_{xz}\right)\end{array}\right\} \tag{1.8}$$

Corresponding Author: Dr. Muhammad Qaiser Shahbaz, Department of Mathematics, COMSATS Institute of Information Technology, Lahore, Pakistan

Samiuddin and Hanif [4] have given the following regression estimator in two phase sampling using two auxiliary variables:

$$\bar{y}_{SH(2)} = \bar{y}_2 + \alpha\left(\bar{x}_1 - \bar{x}_2\right) + \beta\left(\bar{z}_1 - \bar{z}_2\right) \qquad (1.9)$$

The variance of (1.9) given by Samiuddin and Hanif [4] is:

$$\mathrm{Var}\left(\bar{y}_{SH(2)}\right) = S_y^2\left\{\theta_2\left(1-\rho_{y.xz}^2\right) + \theta_1\rho_{y.xz}^2\right\} \qquad (1.10)$$

where $\rho_{y.xz}$ is multiple correlation coefficient between *Y* and combined effects of *X* and *Z*.

Several other estimators have been proposed for estimation of population characteristics in survey sampling. In the following section we have proposed a general class of shrinkage estimators in survey sampling.

MAIN RESULT

In this section we have proposed a general class of shrinkage estimator for estimation of any population characteristic. We proposed the estimator alongside its MSE in the following theorem.

Theorem 1: The general class of shrinkage estimator in survey sampling is:

$$\hat{t}_S = \frac{\hat{t}}{1 + T^{-2}\mathrm{MSE}\left(\hat{t}\right)} \qquad (2.1)$$

where $\hat{t}$ is any available estimator of parameter *T*. The minimum MSE of (2.1) is given as:

$$\mathrm{mse}\left(\hat{t}_S\right) = \frac{\mathrm{MSE}\left(\hat{t}\right)}{1 + T^{-2}\mathrm{MSE}\left(\hat{t}\right)} \qquad (2.2)$$

Proof: Suppose an estimator $\hat{t}$ of parameter *T* is available. Let the shrinkage estimator be $\hat{t}_S = d\hat{t}$, where the constant d is obtained by minimizing the MSE of $\hat{t}_S$. The MSE of $\hat{t}_S$ is

$$\mathrm{mse}\left(\hat{t}_S\right) = E\left(\hat{t}_S - T\right)^2 = E\left(d\hat{t} - T\right)^2$$

Suppose further that $\hat{t} = T + \bar{e}_T$ with $E\left(\bar{e}_T\right) = 0$ and

$$E\left(\bar{e}_T^2\right) = \mathrm{Var}\left(\bar{e}_T^2\right) = \mathrm{MSE}\left(\hat{t}\right)$$

Now, using $\hat{t} = T + \bar{e}_T$, the MSE of $\hat{t}_S$ is given as:

$$\mathrm{mse}\left(\hat{t}_S\right) = E\left(d\hat{t} - T\right)^2 = E\left\{d\left(T + \bar{e}_T\right) - T\right\}^2 \qquad (2.3)$$

Partially differentiating (2.3) w.r.t. d and equating to zero we have:

$$E\left[\left\{d\left(T + \bar{e}_T\right) - T\right\}\left(T + \bar{e}_T\right)\right] = 0 \qquad (2.4)$$

Applying expectation and simplifying we have:

$$d = \left\{1 + T^{-2}\mathrm{mse}\left(\hat{t}\right)\right\}^{-1} \qquad (2.5)$$

using (2.5) in $\hat{t}_S = d\hat{t}$, we have (2.1). Now the MSE of $\hat{\theta}_S$ is obtained as under:

$$\mathrm{mse}\left(\hat{t}_S\right) = E\left[\left\{d\left(T + \bar{e}_T\right) - T\right\}\left\{d\left(T + \bar{e}_T\right) - T\right\}\right]$$

Using (2.4) in above equation we have $\mathrm{mse}\left(\hat{t}_S\right) = T^2\left(1 - d\right)$. Further by using (2.5), we have (2.2).

The MSE of $\hat{t}_S$ given in (2.2) clearly indicates that the shrinkage estimator will always outperform the conventional estimator. The shrinkage estimator has one deficiency that it requires the value of *T* which is generally unknown. A consistent shrinkage estimator of parameter *T* can be defined as:

$$\hat{t}_S = \hat{t}\left\{1 + \hat{t}^{-2}\,\widehat{\mathrm{mse}}\left(\hat{t}\right)\right\}^{-1} \qquad (2.5)$$

The estimator of MSE of $\hat{t}_S$ is given as:

$$\widehat{\mathrm{mse}}\left(\hat{t}_S\right) = \widehat{\mathrm{mse}}\left(\hat{t}\right)\left\{1 + \hat{t}^{-2}\,\widehat{\mathrm{mse}}\left(\hat{t}\right)\right\}^{-1} \qquad (2.6)$$

The equations (2.5) and (2.6) can be used to construct the confidence interval for the parameter *T*.

In the following we have given some examples of the shrinkage estimator (2.1) for estimation of the population parameters.

APPLICATIONS

The shrinkage estimator has been proposed in (2.1). In this section we have given some applications of (2.1).

Example 1: Shrinkage Mean per Unit Estimator can be obtained by using $\hat{t} = \bar{y}$ in (2.1). The estimator is given as:

$$\bar{y}_S = \bar{y}\left\{1 + \theta \bar{Y}^{-2} S_y^2\right\}^{-1} \tag{3.1}$$

The MSE of (3.1) can be obtained by using (1.2) in (2.2) which give the following expression for MSE of (3.1):

$$\text{mse}\left(\bar{y}_S\right) = \theta S_y^2 \left\{1 + \theta \bar{Y}^{-2} S_y^2\right\}^{-1} \tag{3.2}$$

Clearly $\text{mse}\left(\bar{y}_S\right) < \text{Var}\left(\bar{y}\right)$. The consistent shrinkage estimator is given as:

$$\bar{y}_S = \bar{y}\left\{1 + \theta \bar{y}^{-2} s_y^2\right\}^{-1} \tag{3.3}$$

With an estimator of MSE as:

$$\widehat{\text{mse}}\left(\bar{y}_S\right) = \theta s_y^2 \left\{1 + \theta \bar{y}^{-2} s_y^2\right\}^{-1} \tag{3.4}$$

Example 2: Shrinkage Regression Estimator can be obtained by using $\hat{t} = \bar{y}_{lr}$ in (2.1). In this construction we assume that $\bar{y}_{lr}$ is almost unbiased estimator. The estimator is given as:

$$\bar{y}_{lrS} = \left\{\bar{y} + \beta\left(\bar{X} - \bar{x}\right)\right\}\left\{1 + \theta \bar{Y}^{-2} S_y^2 \left(1 - \rho^2\right)\right\}^{-1} \tag{3.5}$$

The MSE of (3.5) is given as:

$$\text{mse}\left(\bar{y}_{lrS}\right) = \theta S_y^2 \left(1 - \rho^2\right)\left\{1 + \bar{Y}^{-2} \theta S_y^2 \left(1 - \rho^2\right)\right\}^{-1} \tag{3.6}$$

Comparing (1.4) with (3.6) we can see that $\text{mse}\left(\bar{y}_{lrS}\right) < \text{Var}\left(\bar{y}_{lr}\right)$.

The consistent estimator can be easily obtained by replacing population parameters with corresponding estimates.

Example 3: Shrinkage Two Phase Sampling Regression Estimator is obtained by using $\hat{t} = \bar{y}_{lr(2)}$ in (2.1). The shrinkage estimator is given as:

$$\bar{y}_{l(2)S} = \left\{\bar{y}_2 + \beta\left(\bar{x}_1 - \bar{x}_2\right)\right\} \left[1 + \bar{Y}^{-2} S_y^2 \left\{\theta_2\left(1 - \rho^2\right) + \theta_1 \rho^2\right\}\right]^{-1} \tag{3.7}$$

The MSE of $\bar{y}_{l(2)S}$ is given as:

$$\text{mse}\left(\bar{y}_{l(2)S}\right) = \frac{S_y^2\left\{\theta_2\left(1 - \rho^2\right) + \theta_1 \rho^2\right\}}{1 + \bar{Y}^{-2} S_y^2 \left\{\theta_2\left(1 - \rho^2\right) + \theta_1 \rho^2\right\}} \tag{3.8}$$

Clearly $\text{mse}\left(\bar{y}_{l(2)S}\right) < \text{Var}\left(\bar{y}_{lr(2)}\right)$. The consistent shrinkage regression estimator for two phase sampling is given as:

$$\bar{y}_{l(2)S} = \left\{\bar{y}_2 + b_{yx}\left(\bar{x}_1 - \bar{x}_2\right)\right\} \left[1 + \bar{y}^{-2} s_y^2 \left\{\theta_2\left(1 - r^2\right) + \theta_1 r^2\right\}\right]^{-1} \tag{3.9}$$

where r is sample correlation coefficient between X and Y. The estimate of MSE can be obtained by using sample estimates in (3.8).

Example 4: Shrinkage Multiple Regression Estimator can be obtained by using

$$\hat{t} = \bar{y}_{mlr} = \bar{y} + \sum_{j=1}^{p} \beta_j\left(\bar{X}_j - \bar{x}_j\right)$$

in (2.1). The resulting shrinkage estimator is given as:

$$\bar{y}_{mlrS} = \bar{y}_{mlr}\left\{1 + \theta \bar{Y}^{-2} S_y^2 \left(1 - \rho^2_{y.x_1x_2..x_p}\right)\right\}^{-1} \tag{3.10}$$

The MSE of (3.10) is given as:

$$\text{MSE}\left(\bar{y}_{mlrS}\right) = \left\{\theta S_y^2 \left(1 - \rho^2_{y.x_1x_2..x_p}\right)\right\} \left\{1 + \theta \bar{Y}^{-2} S_y^2 \left(1 - \rho^2_{y.x_1x_2..x_p}\right)\right\}^{-1} \tag{3.11}$$

The consistent estimator is given as:

$$\bar{y}_{mlrS} = \bar{y}_{mlr}\left\{1 + \theta \bar{y}^{-2} s_y^2 \left(1 - R^2_{y.x_1x_2..x_p}\right)\right\}^{-1} \tag{3.12}$$

where $R^2_{y.x_1x_2..x_p}$ is sample multiple correlation coefficient between Y and the combined effect of all the auxiliary variables.

Again the shrinkage version of multiple regression estimator perform better as compared with the conventional multiple regression estimator.

From above applications we can see that the general shrinkage estimator can be used to develop several estimators.

REFERENCES

1. Cochran, W.G., 1977. Sampling Techniques. J. Wiley New York.
2. Hansen, M.H., W.N. Hurwitz and W.G. Madow, 1953. Sample Survey Methods and Theory. John Wiley, New York, Vol: 2.
3. Kiregyera, B., 1984. Regression-type estimator using two auxiliary variables and model of double sampling from finite populations. Metrika, 31: 215-226.
4. Samiuddin, M. and M. Hanif, 2007. Estimation of population mean in single and two phase sampling with or without additional information. Pak. J. Stat., 23 (2): 99-118.