STUDIES ON SOME ESTIMATORS IN DOUBLE SAMPLING

ABSTRACT

GITASREE DAS

DEPARTMENT OF ECONOMICS
SCHOOL OF SOCIAL SCIENCES

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENT OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

NORTH-EASTERN HILL UNIVERSITY

SHILLONG, INDIA
AUGUST, 1991

It is a well known fact that for estimating the population mean μ_y of the random variable Y, precision of the estimator can be increased when information on an auxiliary variable X, highly correlated with Y is readily available on all the units of the population, incorporating the knowledge of μ_x, the population mean of X. When the relationship between Y and X is found to be approximately linear but the line does not go through the origin, linear regression estimate may be used. To use the linear regression estimator it is usually assumed that the population mean μ_x is known. However, in certain practical situations μ_x is not known a priori, in which case the technique of double sampling is applied. Here one may take a preliminary sample to estimate it. And a linear regression estimator can be defined utilising this estimated mean of the auxiliary variable.

In certain situations, the experimenter may have partial information about μ_x. Han (1973) has suggested the use of double sampling with partial information on the auxiliary variable. In order to utilise the partial information one can perform a preliminary test about the hypothesis that $\mu_x = \mu_o$, where μ_o is the value obtained from the partial information. After the preliminary sample is obtained, he can test $H_o: \mu_x = \mu_o$ against $H_1: \mu_x \neq \mu_o$. If H_o is accepted μ_o will be used in the regression estimator; if H_o is rejected, the sample mean based on the preliminary sample is used. This estimator is usually called the preliminary test estimator.

In estimating the population mean μ_y of the random variable Y, suppose that in addition to information on an auxiliary variable X, information on yet another auxiliary variable Z is available. When μ_x is not known, one can take a preliminary sample of size n' to estimate it. Again if μ_z is also not known, assume that Z is known over another large sample, also of size n'. In such a situation an estimator t_3 using X and Z is being suggested by Mukherjee, Rao and Vijayan (1987).

When μ_z is unknown and the experimenter has partial information about it, a preliminary test $H_0: \mu_z = 0$ (letting $\mu_0 = 0$ without loss of generality) can be employed. If H_0 is accepted, μ_0 will be used in the regression estimator; if H_0 is rejected, the sample mean based on the preliminary sample consisting of n' independent observations on Z is used. Since μ_x is unknown, it is estimated from another preliminary sample also of size n'. In such a situation we suggested a preliminary test estimator, t_4, in double sampling with two auxiliary variables having partial information on one auxiliary variable. The bias function of t_4 was obtained and it was observed that when μ_z increases from 0, Bias (t_4) increases to a maximum, then decreases to zero. The bias is very close to zero at $\mu_z = 1$. The bias found here is quite small almost in all cases.

Next, we tried to obtain the mean square error of t_4. It was observed that the relative efficiency e_4 of t_4 with respect to t_3 has a maximum at $\mu_z = 0$; when μ_z increases e_4 decreases to a minimum and then increases to unity. It is found that e_4 is very close to 1 at $\mu_z = 1$. Optimum allocation for sample sizes were made for the estimator t_4, for a given cost function. And we obtained the conditions under which $M_{4,opt} \leqslant V_{opt}(t_3)$ with equality holding for $Z_\alpha = 0$, which is the case when the two estimators coincide.

Another estimator t_5 was suggested, which is also a preliminary test estimator under double sampling. However, unlike before, in this case we have partial information on both the auxiliary variables. A bias function was obtained and in this case also it was observed that Bias $(t_5) = 0$ at $\mu_x = \mu_z = 0$. Also as μ_x, μ_z increase from 0, Bias (t_5) increases to a maximum and again decreases to zero. Bias is very close to zero at $\mu_x = \mu_z = 1$. Almost in all cases the bias is very negligible.

Mean square error of t_5 and its relative efficiency with respect to t_3 was obtained. For a given cost function optimum allocation for sample sizes were made and we proved that $M_{5,opt} \leqslant V_{opt}(t_3)$, with equality holding for $Z_\alpha = 0$, when the two estimators coincide. We had earlier obtained a similar result for t_4 as well. However, we have further shown that $MSE(t_5)$ is even smaller than

$MSE(t_4)$. Next, we proved that $MSE(t_5) \leqslant MSE(t_2)$, t_2 being the preliminary test estimator with one auxiliary variable suggested by Han (1973).

Next we suggested an estimator t_7 of more generalized nature in the sense that the preliminary samples for the two auxiliary variables X and Z are now taken to be of unequal sizes, which was not so in the earlier cases. However, even in this case we find that the function, Bias (t_7) behaves in the manner similar as before. Here also we obtained the $MSE(t_7)$ and relative efficiency of t_7. The optimum allocation of the three sample sizes, viz., n, n' and n" were made under a given linear cost function. Again, we have also proved that under the optimum conditions MSE $(t_7) \leqslant V(t_6)$, where t_6 is the regression-type estimator in double sampling with two auxiliary variables having preliminary samples of unequal sizes without using any preliminary test.

Finally, we did some empirical studies in which different data sets were considered to demonstrate the practical use of different estimation formulae and empirically demonstrate the performence of the suggested estimators. In all the cases, under the stated assumptions, satisfactory results were obtained. Finally we conclude that under double sampling preliminary test estimators are more efficient than the usual regression -type estimators.