International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 1, 2002, pp 62–64

A Technical Report On Multivariate Regression-Type Estimators

D.C. Roy
Department of Statistics, University of Rajshahi, Bangladesh.

M.C. Agrawal
Department of Statistics, University of Delhi, India

[Received June 8, 2002; Revised October 11, 2001; Accepted December 12, 2002]

In this report, we rationalize and reconcile the sources of divergence in mean square errors of the three multivariate regression-type estimators arrived at by Mukerjee et al. (1987) and Ahmed (1998).

Let y and x denote the study and the auxiliary variables respectively taking values y_i for the ith $(i = 1, 2, \cdots, N)$ unit of a finite population. The problem of estimating the population mean $\bar{Y}_N$ of y-variable when the population mean $\bar{X}_N$ of x variable is known has been dealt with in literature quite extensively. However, in certain practical situations when $\bar{X}_N$ is not known a priori, the technique of two-phase sampling is effectively exploited. This sampling procedure requires collection of information on x for the first-phase sample s' of size $n'(n' < N)$ and on y for the second-phase sample of size $n(n < n')$ selected from the first-phase sample s'.

Sometimes, even if $\bar{X}_N$ is unknown, information on a second auxiliary variable z is available on all units of the population and let $\bar{Z}_N$ be the population mean of z-variable. Kiregyera (1984) and Mukerjee et al. (1987), exploiting this idea, suggested certain regression-type estimators. Mukerjee et al. (1987), who suggested three regression-type estimators, also presented a corrected form of the mean square error (MSE) of an estimator due to Kiregyera (1984). Ahmed (1998) found fault with the MSE expression for each of the three estimators mooted by Mukerjee et al. (1987) and offered the 'corrected' forms which led him to conclude that the estimators due to Mukerjee et al. (1987) are, contrary to their findings, no more better than the one due to Kiregyera (1984). He also noted that a certain estimator due to Tripathi and Ahmed (1995) is the best among the potentially competing estimators.

We observed that the MSE expressions as obtained by Mukerjee et al. (1987) are correct if the regression coefficients involved in the three regression-type estimators are appropriately interpreted. In other words, a meaningful explanation ascribed to the regression coefficients involved in the regression-type estimators due to Mukerjee et al. (1987) renders the comments and findings of Ahmed (1998) untenable.

Kiregyra (1984) has considered the following estimator using information on both x and z

$$t_2 = \bar{y}_n + b_{yx}[(\bar{x}'_n - \bar{x}_n) - b_{xz}(\bar{z}'_n - \bar{Z}_N)]$$

where b_{yx} and b_{xz} are regression coefficients of y on x and of x on z respectively based on the largest possible sample. For large N, Mukerjee et al. (1987), to terms of $O\left(\frac{1}{n}\right)$, corrected the MSE of t_2 as

$$\mathrm{MSE}(t_2) = \sigma_y^2 \left[\frac{1-\rho_{yx}^2}{n} + \frac{\rho_{yx}^2 + \rho_{yx}^2\rho_{xz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}}{n'}\right]$$

where ρ_{yx}, ρ_{yz} and ρ_{xz} are the simple correlation coefficients of y and x, y and z and x and z respectively and $\sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^2$.

Mukerjee et al. (1987) have also suggested the three estimators, viz.

$$\begin{aligned}
t_3 &= \bar{y}_n + b_{yx}(\bar{x}'_n - \bar{x}_n) + b_{yz}(\bar{z}'_n - \bar{z}_n) \\
t_4 &= \bar{y}_n + b_{yx}(\bar{x}'_n - \bar{x}_n) + b_{yz}(\bar{Z}_N - \bar{z}_n) \\
t_5 &= \bar{y}_n + b_{yx}(\bar{x}'_n - \bar{x}_n) + b_{yx}b_{xz}(\bar{Z}_N - \bar{z}'_n) + b_{yz}(\bar{Z}_N - \bar{z}_n)
\end{aligned}$$

wherein the sample regression coefficients, b_{yx}, b_{yz} and b_{xz} have been prima facie interpreted by Ahmed (1998) as the ordinary ones, meaning thereby that b_{yx}, b_{yz} and b_{xz} are the regression coefficients of y on x, of y on z and of x on z respectively, the first two being based on the sample of size n and the third one being based on the sample of size n'. Driven by this consideration, Ahmed (1998) obtained the MSE's of t_3, t_4 and t_5 as

$$\begin{aligned}
\mathrm{MSE}(t_3) &= \sigma_y^2\left[\frac{1-(1-\rho_{yx}^2)\rho_{y.xz}^2}{n} + \frac{(1-\rho_{xz}^2)\rho_{y.xz}^2}{n'}\right] = V_3, \text{ say} \\
\mathrm{MSE}(t_4) &= \sigma_y^2\left[\frac{1-(1-\rho_{xz}^2)\rho_{y.xz}^2}{n} + \frac{(1-\rho_{xz}^2)\rho_{y.xz}^2 - \rho_{yz}^2}{n'}\right] = V_4, \text{ say} \\
\mathrm{MSE}(t_5) &= \sigma_y^2\left[\frac{1-(1-\rho_{xz}^2)\rho_{y.xz}^2}{n} + \frac{\rho_{yx}^2 + \rho_{yx}^2\rho_{xz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}}{n'}\right] = V_5, \text{ say.}
\end{aligned}$$

However, if we look upon the regression coefficients b_{yx} and b_{yz} involved in t_3, t_4 and t_5 as partial regression coefficients and denote them for the sake of clarity by $b_{yx.z}$ and $b_{yz.x}$ respectively, then we can designate these estimators more explicitly as

$$\begin{aligned}
t_3^* &= \bar{y}_n + b_{yx.z}(\bar{x}'_n - \bar{x}_n) + b_{yz.x}(\bar{z}'_n - \bar{z}_n) \\
t_4^* &= \bar{y}_n + b_{yx.z}(\bar{x}'_n - \bar{x}_n) + b_{yz.x}(\bar{Z}_N - \bar{z}_n) \\
t_5^* &= \bar{y}_n + b_{yx.z}(\bar{x}'_n - \bar{x}_n) + b_{yx.z}b_{xz}(\bar{Z}_N - \bar{z}'_n) + b_{yz.x}(\bar{Z}_N - \bar{z}_n).
\end{aligned}$$

The MSE's of t_3^*, t_4^* and t_5^* will be the same as obtained by Mukerjee et al. (1987), viz.

$$\begin{aligned}
\mathrm{MSE}(t_3^*) &= \sigma_y^2\left[\frac{1-\rho_{y.xz}^2}{n}+\frac{\rho_{y.xz}^2}{n'}\right]=V_3^*,\ \text{say}\\
\mathrm{MSE}(t_4^*) &= \sigma_y^2\left[\frac{1-\rho_{y.xz}^2}{n}+\frac{(\rho_{yx}-\rho_{yz}\rho_{xz})^2}{n'(1-\rho_{xz})^2}\right]=V_4^*,\ \text{say}\\
\mathrm{MSE}(t_5^*) &= \sigma_y^2\left[\frac{1-\rho_{y.xz}^2}{n}+\frac{(1-\rho_{yz}^2)\rho_{xy.z}^2}{n'}\right]=V_5^*,\ \text{say}
\end{aligned}$$

where $\rho_{y.xz}$ and $\rho_{xy.z}$ are the multiple and the partial correlation coefficients, given by

$$\begin{aligned}
\rho_{y.xz}^2 &= \frac{\rho_{yx}^2+\rho_{yz}^2-2\rho_{yx}\rho_{yz}\rho_{xz}}{(1-\rho_{xz}^2)}\\
\rho_{xy.z}^2 &= \frac{(\rho_{yx}-\rho_{yz}\rho_{xz})^2}{(1-\rho_{xz}^2)(1-\rho_{yz}^2)}.
\end{aligned}$$

In the aforesaid context, it goes without denying that the symbols (representing the regression coefficients) employed by Mukerjee et al. (1987) were deficient lacking in necessary explanation which they have agreed in a rejoinder (Mukerjee et al.(2000)). This, in fact, left some scope for confusion, thus giving rise to the paper by Ahmed (1998). Finally, we demonstrate that the paper by Ahmed(1998) does not carry any positive findings.

Acknowledgement

The authors are thankful to the referee for valuable suggestions.

References

Ahmed, M.S. (1998). A note on regression-type estimators using multiple auxiliary information. *Austral. & New Zealand J. Statist.*, 40(3), 373 - 376.

Kiregyera, B. (1984). Regression-type estimators using two auxiliary variables and the model of double sampling from finite population. *Metrika*, 31, 215 - 226.

Mukerjee, R., Rao, T. J. & Vijayan, K. (1987). Regression-type estimator using multiple auxiliary information. *Austral. J. Statist.*, 29, 244 - 254.

Mukerjee, R., Rao, T. J. & Vijayan, K. (2000). Rejoinder to 'Ahmed, M. S.(1998) - A note on regression-type estimators using multiple auxiliary information.' *Austral. & New Zealand J. Statist.*, 42(2), 245.

Tripathi, T. P. & Ahmed, M. S. (1995). A class of estimators for a finite population based on multivariate information and general two-phase sampling. *Calcutta Statist. Assoc. Bull.*, 45, 203 - 218.