Sankhyā : The Indian Journal of Statistics
1993, Volume 55, Series B, Pt. 2, pp. 141–155

ESTIMATION OF A FINITE POPULATION TOTAL UNDER REGRESSION MODELS : A REVIEW*

By P. MUKHOPADHYAY

Indian Statistical Institute

SUMMARY. Estimation of a finite population total under prediction approach using regression superpopulation models has engaged the attention of survey statisticians over more than the last two decades. Some of these model-dependent and model-based investigations have been reviewed in this article. It has also been shown that a sampling design-based conventional strategy fares better than some optimal model-dependent procedures on an average from the point of view of robustness.

0. Introduction

This paper makes a review of some works on estimating a finite population total based on the assumption of an underlying superpopulation model.

1. Model dependent optimal prediction

We denote by $\mathcal{P}$ a finite population of N identifiable units $\{1, ..., i, ..., N\}$; 'y_i' value of 'y' (character of interest) on i, p a fixed size (n) sampling design ($s.\ d$) with inclusion-probabilities π_i, $\pi_{ii'}$ used toes timate the population total $Y(= \Sigma\, y_i)$ by choosing a sample (set) s with probability $p(s)$, $\bar{s} = \mathcal{P} - s$, $\mu_n = \{p\}$. Under prediction theory based approach, $\boldsymbol{y} = (y_1, ..., y_N{}')'$ is considered as a realisation of $\boldsymbol{Y} = (Y_1, ..., Y_N)'$ [Y_i being a r.v. having a value y_i] having a joint distribution $\eta_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} = (\theta_1, ..., \theta_p) \epsilon\, H$(parameter space). $\eta_{\boldsymbol{\theta}}$ may belong to a class $C = \{\eta_{\boldsymbol{\theta}}\}$, called the superpopulation model. Given the data $d = \{(k, y_k), k \epsilon s\}$ one draws inference about $Y = \Sigma\, Y_i$ (now a r.v) on the basis of prior η, using a predictor $\hat{T}_s$. $\hat{T}_s$ is m-(model-) unbiased if $\mathcal{E}(\hat{T}_s - Y) = 0 \;\forall\; s : p(s) > 0$ andpm -(design-model-) unbiased if $E\, \mathcal{E}(\hat{T}_s - Y) = 0$. E, V, $\mathcal{E}$, $\mathcal{V}$, $\mathcal{C}$ will denote respectively p—(design-) expectation, p-variance, m-expectation, m-variance, m-covariance. For a non-informative design ($p(s)$ is independent of $\boldsymbol{y}$), order of operations E, $\mathcal{E}$ are interchangeble.

Paper received- February 1991.

AMS (1990) *subject classification.* 62 D05.

Keywords and phrases. Polynomial (multiple) regression model, BLU—predictor, purposive (balanced) sampling design, asymptotic design unbiasedness (consistency), asymptotic mean square error.

*Paper prepared for the International Conference, 'Recent Developments in Probability and Statistics', New Delhi, December 1990, dedicated to the memory of Professor C. G. Khatri.

The problem studied in this section were foreshadowed and considered by Brewer (1963). Royall (1970, 1971), Royall and Herson (1973) gave general formulations to the problem in their elegant papers.

Since (Royall and Herson, 1973) $Y = \sum_s Y_k + \sum_{\bar{s}} Y_k$, after d have been collected, $Y = \sum y_k + \sum_{\bar{s}} Y_k$ and hence

$$\hat{T}_s = \sum_s y_k + \hat{U}_s,$$

where $\hat{U}_s$ is a predictor of $\sum_{\bar{s}} Y_k$. $\hat{T}_s$ is m-unbiased for Y if $\hat{U}_s$ is so for $\sum_{\bar{s}} Y_k$.

Two types of mean square errors (mse) of strategy $(p, \hat{T}_s)$ of interest are : (*i*) $\mathcal{E}\, E\,(\hat{T}_s - Y)^2 = M\,(p, \hat{T})$ (say) (*ii*) $E\,\mathcal{E}(\hat{T}_s - \mu)^2 = M_1\,(p, \hat{T})$ (say), where $\mu = \mathcal{E}(Y)$. It has been recommended (Sarndal, 1980) that when one's real interest is in the present population from which the sample has been drawn, one should use M for choosing an optimal strategy. For deriving an optimal present predictor of Y for some future population which is of the same type as the survey population (having the same μ), one's real interest is actually in μ and here M_1 should be used. Relation between M and M_1 is

$$M(p, \hat{T}) = E\mathcal{V}(\hat{T}) + E(\mathcal{B}(\hat{T}))^2 + \mathcal{V}(Y) - 2\mathcal{E}\{(Y-\mu)E(\hat{T}-\mu)\} \qquad \ldots \quad (1.1)$$

where $\mathcal{B}(\hat{T}) = \mathcal{E}\,(\hat{T} - Y)$, m-bias in T.

For a m-unbiased $\hat{T}$,

$$M(p, \hat{T}) = E\left[\mathcal{V}(\hat{U}_s) + \mathcal{V}\left(\sum_{\bar{s}} Y_k\right) - 2\,\mathcal{C}\left(\hat{U}_s, \sum_{\bar{s}} Y_k\right)\right] \qquad \ldots \quad (1.2)$$

If η is a product-measure (Y_i's are independent) M is minimum if $\mathcal{V}(\hat{U}_s)$ is minimum and thus for a given s the optimal m-unbiased predictor of Y is

$$\hat{T}_s^+ = \sum_s y_k + \hat{U}_s^+$$

where $\hat{U}_s^+$ is minimum (m–) variance (m–) unbiased predictor of $\sum_{\bar{s}} Y_k$. An optimal strategy $(p^+, \hat{T}^+)$ in the class $(\bar{\mu}, \tau)$ is, therefore, one for which

$$M(p^+, \hat{T}^+) \leqslant M(p, \hat{T}) \quad \forall\, p\, \epsilon \bar{\mu},\, \hat{T}\, \epsilon\, \tau. \qquad \ldots \quad (1.3)$$

The form of $\hat{T}^+$ does not depend on the s.d. (unlike p-based estimators, say, Horvitz-Thompson estimator, e_{HT}). After $\hat{T}^+$ is obtained, p^+ is chosen through

(1.3). The emphasis on this m-dependent approach is, therefore, on a correct postulation of η and generating $\hat{T}_S$. The choice of a suitable p takes a secondary role.

Optimal prediction under some models: polynomial regression model $\eta(\delta_0, \ldots, \delta_J ; v(x))$. Assume x_i, value of an auxiliary variable x on i is known $(i = 1, \ldots, N)$, Y_i, $Y_{i'}$, $(i \neq i')$ are uncorrelated with

$$\mathcal{E}(Y_i \mid x_i) = \sum_0^J \delta_j \beta_j x_i^j$$
$$\mathcal{V}(Y_i \mid x_i) = \sigma^2 v(x_i), \qquad \ldots \quad (1.4)$$

β_j's, σ^2 are unknown constants, $v(x_i)$, a known function of x_i, $\delta_i = 1(0)$ if β_j is present (absent) in $\mathcal{E}(Y_i)$. The best linear (m–) unbiased predictor (BLUP) of Y, for a given s, is

$$\hat{T}^*_s(\delta_0, \ldots, \delta_J ; v(x)) = \sum_s y_k + \sum_0^J \delta_j \hat{\beta}^*_j \sum_{\bar{s}} x_k^j$$

where $\hat{\beta}^*_j$ is the BLUP of β_j under η (obtained by Gauss-Markoff theorem). Royall (1970) studied $\hat{T}^*(0, \ 1 ; x^g) = \hat{T}^*_g(\text{say})$, $g = 0, \ 1, 2$ $\left(\hat{T}^*_1\right.$ is the ratio predictor $\left.\hat{T}_R = \frac{\overline{Y}_s}{\bar{x}_s} X\right)$ and proved.

Theorem 1.1. *If (a) $v(x)$ is non-decreasing function of x, (b) $v(x)/x^2$ is a non-increasing function of x, optimal s.d. to use $\hat{T}^*_g$ in μ_n is the purposive design p^* where the sample s^* having units with the highest n values of x_i are selected with centainty.*

Multiple regression models. (a) Y_i, $Y_{i'}$, $(i \neq i')$ uncorrelated: Assume for each i, values x_{ij} of $(r+1)$ auxiliary variables $x_j (j = 0, 1, \ldots, r)$ are known and Y_i, $Y_{i'}$ $(i \neq i')$ are uncorrelated with

$$\mathcal{E}(Y_i \mid \boldsymbol{x}_i') = \boldsymbol{x}_i' \boldsymbol{\beta}$$
$$\mathcal{V}(Y_i \mid \boldsymbol{x}_i) = \sigma^2 v(\boldsymbol{x}_i) = \sigma^2 v_i \qquad \ldots \quad (1.5)$$

where $\boldsymbol{x}_i' = (x_{i0}, \ldots, x_{ir})$, $\boldsymbol{\beta}' = (\beta_0, \ldots, \beta_r)$. Often (as is here), $x_{i0} = 1 \ \forall i$. Denoting $\boldsymbol{Y}_s = [Y_k, k \in s]_{n \times 1} \boldsymbol{Y}_{\bar{s}} = [Y_k, k \in \bar{s}]_{(N-n) \times 1}$, $\boldsymbol{X} = ((x_{ij}, \ i = 1, \ldots, N;$ $j = 0, \ldots, r))_{N \times (r+1)}$ $\boldsymbol{X}_s = ((x_{ij}, \ \ i \in s, j = 0, \ \ldots, \ \ r))_{n \times (r+1)}$, $\boldsymbol{X}_{\bar{s}} = ((x_{ij}, i \in \bar{s},$ $j = 0, \ldots, r)_{(N-n) \times (r+1)}$, $\boldsymbol{1}_p' = (1, \ldots, 1)_{p \times 1}$, $x_s' = (x_{s0}, \ldots, x_{rs})$, $\bar{\boldsymbol{x}}_s' = (\bar{\boldsymbol{x}}_{0s}, \ldots,$ $\bar{\boldsymbol{x}}_{rs})$, $\bar{\boldsymbol{x}}_{js} = \sum_s x_{ij}/n = \boldsymbol{x}_{js}/n$ (and $\boldsymbol{x}_{\bar{s}}'$), $\bar{\boldsymbol{x}}_{\bar{s}}'$ similarly),

$V = \text{Diag}\ (v_k\ ;\ k = 1, \ldots, N)$, $Y_s = \text{Diag}\ (v_k\ ;\ k \in s)$ (and $V_{\bar{s}}$ similarly), and assuming wlg $Y' = [Y_s'), Y_{\bar{s}}']$, the model (1.5), denoted as $\eta(X, v)$ is

$$\mathcal{E}(Y) = X\,\beta,\ \mathcal{D}(Y) = \sigma^2 V \quad \ldots \quad (1.6)$$

($\mathcal{D}(\cdot)$) denoting the dispersion matrix). The BLUP of Y is Royall (1975)

$$\hat{T}^*(X, v) = \sum_s y_k + x_{\bar{s}}'\,\hat{\beta}^* \quad \ldots \quad (1.7)$$

where $$\hat{\beta}^* = (X_s'\,V_s^{-1}\,X_s)^{-1}(X_s\,V_s^{-1}Y_s).$$

(b) Y_i, $Y_{i'}$ not all uncorrelated : Denoting $\mathcal{C}(Y_{i'}\,Y_{i}^{'}) = \sigma^2\,v_{ii'}'$, $(v_{ii} = v_i)$,

$$\bar{V} = ((v_{ii'}))_{N\times N} = \begin{pmatrix} \bar{V}_s & V_{s,\bar{s}} \\ n\chi n & \\ V_{\bar{s},s} & \bar{V}_{\bar{s}} \end{pmatrix} \quad \ldots \quad (1.8)$$

$K = [\bar{V}_s', V_{s,\bar{s}}]$, the BLUP of Y under this model, denoted as $\eta(X, V)$ is (Royall (1976), Tam (1986))

$$\hat{T}^*(X, V) = 1_n'\,Y_s + 1_{N-n}'\,[X_{\bar{s}}\,\hat{\beta}^* + V_{\bar{s},s}\,\bar{V}_s^{-1}\,(Y_s - X_s\,\hat{\beta}^*)]. \quad \ldots \quad (1.9)$$

In particular when Y_i, Y_i', $(i \neq i')$ have a constant correlation $\rho \in \left[-\dfrac{1}{N-1}, 1\right]$ (denoting this model as $\eta(X, V, \rho)$) we have (Isaki and Fuller, 1982),

Theorem 1.2. *Assume (a)* $\sqrt{v} \in \mathcal{C}(X)$, *where* $\sqrt{v} = (\sqrt{v_1}, \ldots, \sqrt{v_N})'$ *and* $\mathcal{C}(Z)$ *denotes the column-space of* Z ; or (b) $\rho = 0$. *The BLUP of* Y, $\hat{T}^*\ (X, V, \rho)$ *is given by* (1.7).

If, further, Y follows a N-variate normal distribution we have an exchangeable general linear model (EGLM) (Arnold, 1979) and (Mukhopadhyay, 1991),

Theorem 1.3. *Under EGLM,* $\hat{T}^*$ *given in* (1.7) *is UMVU in the class of all unbiased predictors of* Y.

Rodrigues *et al.* (1985) extended the concepts to develop a general theory of prediction which covers both linear and quadratic functions of population values. Skinner (1983) considered the multivariate prediction of mean.

Projection predictors. When predicting the mean of a future population of the same type as the present one, one's real interest is actually in μ and the the optimal prediction is obtained by minimising $M_1\ (p, \hat{T})$

Cochran (1977, p. 159), Sarndal (1980 a, b), Wright (1983), Isaki and Fuller (1982) considered such optimal predictors $\hat{\hat{T}}^*\ (\eta)$ under η. Here all Y_k's (including those with $k \epsilon s$) are predicted and d are used only to predict the model-parameters. Thus

$$\hat{T}^* (X, v) = X \hat{\beta}^*.$$

We have (Sarndal, 1980a),

Theorem 1.4. *Under* η $(X, v), \hat{\hat{T}}^* = \hat{T}^*$ *iff* $\hat{\hat{T}}^*$ *based on the entire* $\mathcal{P}$ *equals* Y.

2. Robustness of $\hat{T}^*$ (η) under alternative models

Royall and Herson (1973) showed that $\hat{T}_R$ remains unbiased under $\eta'(\delta''_0, \ldots, \delta_J\, ;\, v'(x))$ on a balanced sample $s_b(J)$ of order J, which satisfies

$$\bar{x}_s^{(j)} = \bar{X}^{(j)} \quad j = 1, \ldots, J$$

$\bar{x}_s^{(j)} = n^{-1} \Sigma\, x_k^j$, $\bar{X}^{(j)} = N^{-1} \Sigma\, x_k^j$. $\hat{T}^*(\eta)$ will, in general, have non-zero bias $B_{\eta'} \{\hat{T}(\eta)\}$ under η'. For a given s, one should, therefore, choose the one which is least biased (most bias-robust) in a class of rival predictors. The following theorem has been proved (Mukhopadhyay, 1977).

Theorem 2.1. *Under assumptions,*

$$\beta_j \geqslant 0 \;\forall\, j\, ;\, x_k,\, k \,\epsilon\, s \textit{ are not all equal} \qquad \ldots \quad (i)$$

for samples for which

$$(a) \qquad I'_g(0, s) > 0,\, \hat{T}^*_s\,(0, 1\, ;\, x^{g'}) \overset{\eta_1}{\succ} \hat{T}^*_s\,(0, 1\, ;\, x^g),\, g > g' \qquad \ldots \quad (2.1)$$

(b) $\quad I'_g(j, s) \leqslant 0 \;\forall\, j = 2, \ldots, J$ *with at least one* $I_{g'}(j, s) < 0$, $\hat{T}^*_s\,(0, 1\, ;\, x_{g'}) \overset{\eta_2}{\succ} \hat{T}^*_s\,(0, 1\, ;\, x^g), g > g'$ *where* $\hat{T}_s \overset{\xi}{\succ} \hat{\bar{T}}'_s$ *means* T_s *involves less absolute bias than* $\hat{T}'_s$ *under* ξ *and* $\eta_1 = \eta'\,(1, \delta_1\, ;\; v'\,(x)), \eta_2 = \eta'\,(0, \delta'_1, \ldots \delta'_J\, ;$ $v'\,(x))$ *and*

$$I_g(j, s) = \sum_s x_k^{j+1-g} \sum_{\bar{s}} x_k - \sum_z x_k^{2-g} \sum_{\bar{s}} x_k^j.$$

Under p^*, $\hat{T}^*$ $(0, 1\, ;\, x^2)$ *is most bias-robust in the class* $\{\hat{T}^*(0, 1\, ;\, x^g),\, g \,\epsilon\, [0, 2]\}$, *both wrt* η_1 *and* η_2.

The above result gives a basis for a post-sample selection of robust predidictors.

Scott *et al.* (1978), on generalisation of the concept of *RH*-balanced sampling, noted that under generalised balanced samples S^* (J), which satisfies

$$\frac{\sum_{\bar{s}} x_k^j}{\sum_{\bar{s}} x_k} = \frac{\sum_s x_k^{j+1}/v(x_k)}{\sum_s x_k^2/v(x_k)}, j = 0, 1, \ldots, J \qquad \ldots \quad (2.2)$$

$\hat{T}^*(0, 1 ; v(x))$ remains unbiased under $\eta(\delta_0, \ldots, \delta_J ; V(x))$. They studied the case, $v(x) = x^2$, denoting the corresponding $S^*(J)$ as overbalanced samples $S_0(J)$. It has been shown

Theorem 2.2. *For* $S = S^*(J)$, $\hat{T}^*(0, 1 ; v(x))$ *is BLUP under* $\eta(\delta_0, \ldots, \delta_J ;$ $V(x))$ *provided* $V(x) = v(x) \sum_0^J \delta_j \sigma_j^2 x^{j-1}$, σ_j^2*'s being arbitrary constants.*

Cumberland and Royall (1981) defined a π-balanced sample of order J, $S_\pi(J)$ as one for which $\Delta_j(s) = 0, j = 0, \ldots J$ where

$$\Delta_j(s) = \frac{\bar{X}}{n} \sum_s x_i^{j-1} - \bar{X}^{(j)}. \qquad \ldots (2.3)$$

For a πps–*s.d.* ($\pi_i \alpha\, p_i = x_i/X$, $E(\Delta_j(s)) = 0 \;\forall j$). Kott (1986a) showed that under $\eta(0, \delta_1, \delta_2 ; 1)$ the mean of the ratio predictor $\frac{X}{n} \sum_s \frac{Y_i}{x_i}$ coupled with a $p_\pi(2)$–*s.d.* ($p(S_\pi(2)) = 1$) provides a BLU-prediction strategy. Kott (1986b) also considered asymptotically balanced samples. Pfefferman (1984) considered large sample properties of balanced sample.

Pereira and Rodrigues (1983), Pfefferman (1984) examined the questions when $\hat{T}^*(\boldsymbol{X}, v)$ remains unbiased and BLUP under $\eta^*(\boldsymbol{X}^*, v^*)$ where η^* is based on some additional explanatory variables apart from $x_0, \ldots, x_r$ and v^* is a known function of all these variables. Tallis's (1978) result on when $N\bar{Y}_s$ becomes BLUP under η comes as a particular case of their results.

All types of balanced samples are really non-existent in practice. Royall and Herson (1973), Royall and Pfefferman (1982) recommended srs, appropriately stratified random sampling as approximately balanced sampling. Herson (1976) empirically assessed the efficiency of ratio estimator of total using conventional unrestricted random sampling, extreme and balanced sampling plans for sample size ranging from small to large. Cumberland and Royall (1981) proposed the following s.d. to approximate $\bar{p}(1)$ $[\bar{p}(J) : p(S_b(J)) = 1]$:

$$\left[[p(s) = 0 \text{ if } \bar{x}_s - \bar{X} \geqslant \delta, \quad = \frac{1}{c_\delta} \text{ othersise} \right]$$

where δ is a pre-assigned small quantity, c_δ is the number of s with $\bar{x}_s - \bar{X} < \delta$. Iachan (1985) proposed a similar design with some modifications. Mukhopadhyay (1985b) showed srs and pps $\sqrt{x}$ s.d. ($\pi_i \,\alpha \sqrt{x_i}$) are asymptotically equivalent to $\bar{p}(1)$ for using $\hat{T}_R$.

Noting that under η_1, $|B_{\eta_1}\{\hat{T}^*(0, 1 ; x^g)\}| = B_1(g)$ (say) is a monotonically increasing function of g (subject to fulfillment of (i)), Mukhopadhyay (1985a) proposed that for a given s, one can find a g^* for which $B_1(g)$ is near to zero as possible and hence use $\hat{T}^*_s(0, 1 ; x^g)$ as the most bias-robust in the class $\mathcal{J} = \{\hat{T}^*(0, 1 ; x^g),\ g \text{ real}\}$. Similarly $|B_{\eta_2}\{\hat{T}^*(0, 1 ; x^g)\}| = B_2(g)$ (say) being a monotonically decreasing function of g (under (i)), one can find a post-sample most robust predictor $\hat{T}^*_s(0, 1 ; x^{g^{**}})$ in $\mathcal{J}$ under η_2.

3. Model-based predictors

To take care of the brittleness of m-dependent $\hat{T}^*$ and $\hat{\hat{T}}^*$ under departure from the assumed model, the genesis of the model-based predictors which combine both m-randomisation and p-randomisation was evolved. Cassel, Sarndal and Wretman (1976), Sarndal (1980) suggested the generalised regression predictor [GRP] of Y,

$$\hat{T}^*_{GR} = \mathbf{1}_n' \boldsymbol{\pi}_s^{-1}\boldsymbol{Y}_s+(\mathbf{1}' \boldsymbol{X}-\mathbf{1}_n\boldsymbol{\pi}_s^{-1}\boldsymbol{X}_s)\hat{\boldsymbol{\beta}}^* \qquad \dots \quad (3.1)$$

where $\boldsymbol{\pi} = \text{Diag}\ (\pi_k, k = 1, \dots, N)$, $\boldsymbol{\pi}_s = \text{Diag}\ (\pi_k, k \in s)$, $\mathbf{1} = (1, \dots, 1)_{1\times N}$. More generally, the class of GRP of Y is $\hat{T}_{GR}(\boldsymbol{Q})$ obtained by replacing $\hat{\boldsymbol{\beta}}^*$ by

$$\hat{\beta}(\boldsymbol{Q}) = (\boldsymbol{X}_s\boldsymbol{Q}_s\boldsymbol{X}_s)^{-1}(\boldsymbol{X}_s'\boldsymbol{Q}_s\boldsymbol{Y}_s) \qquad \dots \quad (3.2)$$

where $\boldsymbol{Q}$ is an arbitrary positive-definite diagonal matrix of order $N\times N$ and $\boldsymbol{Q}_s$ corresponds to its part in s. $\hat{T}_{GR}(\boldsymbol{Q})$ is m-unbiased under $\eta\ (\boldsymbol{X}, v)$ and is p-biased in general. Noting that a predictor $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is of the form $(\boldsymbol{Z}_s'\boldsymbol{X}_s)^{-1}$ $\boldsymbol{Z}_s'\boldsymbol{Y}_s$, $\boldsymbol{Z}_s = ((Z_{kj}))_{n\times(r+1)}$ being a matrix of suitable weights, Sarndal (1980c) suggested that $\hat{\boldsymbol{\beta}}$ may be called π^{-1} weighted if $\boldsymbol{\pi}_s^{-1}\ \mathbf{1}_s \in \mathcal{C}(\boldsymbol{Z}_s)$. For a π^{-1} weighting in $\hat{\boldsymbol{\beta}}$, $\hat{T}_{GR}$ coincides with the corresponding projction predictor $\hat{T}_{GR}$. Wright (1983), generalising GRP, introduced the $(\boldsymbol{Q}, \boldsymbol{R})$ predictor,

$$\hat{Y}(\boldsymbol{Q}, \boldsymbol{R}) = \mathbf{1}'[\boldsymbol{\Delta R Y}+(\boldsymbol{I}-\boldsymbol{\Delta R})\boldsymbol{X}\hat{\boldsymbol{\beta}}(\boldsymbol{Q})] \qquad \dots \quad (3.3)$$

where $\boldsymbol{\Delta} = \text{Diag}\ (\delta_k, k = 1, \dots, N)$, $\boldsymbol{R} = \text{Diag}\ (r_k, k = 1, \dots, N)$, $\boldsymbol{I} = \text{Diag}$ $(1, \dots, 1)_{N\times N}$, $\delta_k = 1(0)$ if $k \in (\notin) s$. For different choices of $(\boldsymbol{Q}, \boldsymbol{R})$, different types of predictors are obtained. $\boldsymbol{R} = \mathbf{0}$ gives the class of projection predictors, $\boldsymbol{R} = \boldsymbol{I}$, the class of Royall's linear predictors, $\boldsymbol{R} = \boldsymbol{\pi}^{-1}$, the class of GRP.

Sarndal (1980b) had considered a more general strategy. For a s.d. p, his predictor is

$$\hat{T}(p, \boldsymbol{Q}, \boldsymbol{W}_s) = \sum_s w_{ks} Y_k + \sum_0^r \hat{\beta}_j (X_j - \sum_s w_{ks} x_{kj}) \qquad \ldots \quad (3.4)$$

where $\boldsymbol{W}_s$ is a $NX \binom{N}{n}$ matrix of weights w_{ks}, w_{ks} being such that $\sum_s w_{ks} Y_k$ is p-unbiased for Y.

Montanari (1987), generalising Wright's results, defined an enlarged class of $(\boldsymbol{Q}, \boldsymbol{R})$—predictors to include suitable predictors for the model η $(\boldsymbol{X}, \boldsymbol{V})$.

Brewer, Samiuddin, Hanif and Asad (1989) considered a linear p-unbiased estimator $\hat{Y}_E = \sum_s C_{is} y_i$, the weights C_{is} being chosen so as to endow $\hat{Y}_E$ with the ratio estimator property (V $(\hat{Y}_E) = 0$ if $y_i \propto x_i$) and some stability, specially, for outliers. For $n = 2$, they obtained solution to C_{is} by appealing to η (0, 1 ; v_i).

The strategies suggested hare often do not possess any desirable properties (unbiasedness, attainment of a minimum variance bound etc.) in exact analysis, though in asymptotic analysis most of these properties hold.

4. Asymptotically optimum sampling strategies

Brewer (1979) considered the class of predictors which are asymptotically design unbiased (ADU), the predictors being of a particular form suggested by a model and in this class the optimal strategy is one which minimises the asymptotic expected mse (AM). Brewer's stand-point is somewhat between design and pure super-population as a basis for inference. Under his framework for asymptotics, which assumes the repetition of the survey population k times, drawing of independent samples from each (hypothetical) population using the same s.d., producing a pooled predictor of Y based on the combined sample and allowing $k \rightarrow \infty$, the optimal $\hat{T}$, genrated out of $\eta(0, 1\,; \sigma_i^2)$ is

$$\hat{T}^*_{BR} = \sum_s Y_i + \Big\{ \sum_{\bar{s}} Y_i \left(\frac{1}{\pi_i} - 1\right) \Big\{ \Big\{ \sum_s x_i \left(\frac{1}{\pi_i} - 1\right) \Big\}^{-1} \sum_{\bar{s}} x_i \qquad \ldots \quad (4.1)$$

with

$$AM\,(\hat{T}^*_{BR}) = \sum \frac{\sigma_i^2}{\pi_i} - \sum \sigma_i^2 \qquad \ldots \quad (4.2)$$

which equals the minimum value of the average variance of any p-unbiased strategy under the above model, as obtained by Godambe and Joshi (1965), GJLB (say). Rao (1984) obtained Brewer's results under a slightly different approach.

Robinson and Tsui (1981) pointed out some disadvantages in the asymptotic approach considered by Brewer. They considered a sequence of populations $\mathcal{P}_k$ of size $N_{(k)}$ with $N_{(k)} \rightarrow \infty$ as $n_k \rightarrow \infty$, a sample s_k of size $n_{(k)}$ being drawn from $\mathcal{P}_k$ using the similar s.d., with $\lim\limits_{k \rightarrow \infty} \frac{N_{(k)}}{n_{(k)}} < \infty$. Under this framework they proved ADU and ADC of $\hat{T}^*_{BR}$.

Wright, in his asymptotic framework assumed both population size and sample size increasing to infinity with sampling fraction fixed, to conceptualise which he used Brewer's framework and assuming $\eta(\boldsymbol{X}, v)$ showed his $(p, (\boldsymbol{Q}, \boldsymbol{R}))$ strgegy is ADU for $\theta(\boldsymbol{Y}) = \boldsymbol{a}\, \boldsymbol{Y}$ $[\boldsymbol{a}' = (a_1, \ldots, a_N)$, a_i a known constant $(i = 1 \ldots, N)]$ iff

$$(\boldsymbol{I} - \boldsymbol{R}\, \boldsymbol{\pi})\, \boldsymbol{a} = \boldsymbol{Q}\, \boldsymbol{\pi}\, \boldsymbol{X}\, \boldsymbol{\lambda} \qquad \ldots \quad (4.3)$$

for some $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_r)$. He also showed the AM of all the ADU—strategies are given by the GJLB, (4.2). The most efficient stragegy is, threfore, an ADU-predictor with $p : \pi_i \propto a_i \sqrt{v_i}$ when AM is

$$\frac{(\Sigma\, |a_i|\, \sqrt{v_i})^2}{n} - \Sigma\, a_i^2\, v_i = v^* \text{ (say)} \qquad \ldots \quad (4.4)$$

In particular, the class of GRP, $\hat{T}_{GR}$ $(\boldsymbol{Q})$, $\hat{T}^*_{BR}$ along with the corresponding optimal s.d.'s are ADU and asymptotically most efficient.

Wright suggested that when the efficiency of an equal probability ADU—strategy is poor, one can use stratification using *cv* of $|a_i|\, \sqrt{v_i} = c_i$ (say) as stratification variable (*cv* of c_i in stratum $\mathcal{P}_h \leqslant \epsilon \; \forall \; h$) and allocating $n_h \propto \sum\limits_{i \in \mathcal{P}_h} c_i$ and using srs within $\mathcal{P}_h$, when v^* is approximately attained.

Robinson and Sarndal (1983) assumed : (i) there are M nested populations $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_1 \subsetneq \mathcal{P}_2 \subsetneq \ldots$ and $N_1 < N_2 < \ldots$ (ii) a sample s_t of size n_t is selected from $\mathcal{P}_t$ using s.d. p_t with $n_t < n_{t+1}$, $t = 1, 2, \ldots$, samples $s_1, s_2 \ldots$ being not necessarily nested (iii) $M \rightarrow \infty$. Under some regularity conditions on p_t and the moments of the distributions in the model they established ADC and ADU of $\hat{T}_R$ $(\boldsymbol{Q})$ and showed that under $\eta(\boldsymbol{X}, v)$, its AM attains (4.2) and with a suitable design, (4.4).

Fuller and Isaki (1981) showed the AM of the projection predictor

$$\hat{\hat{T}}_{\pi^{-2}} = \boldsymbol{1}'\, \boldsymbol{X}\, \hat{\boldsymbol{\beta}}\, (\boldsymbol{\pi}^{-2})$$

coupled with a πpv—s.d. ($\pi_i \propto v_i$) under $\eta(\boldsymbol{X}, \boldsymbol{V}, \rho)$ is asymptotically minimum in the class of all linear predictors provided $\boldsymbol{v}' \in \mathcal{C}(\boldsymbol{X})$ when $\boldsymbol{v} = (v_1, \ldots, v_N)'$.

Li (1983) considered another type of ADC, which may be termed L_2—consistency. $\hat{T}$ is L_2—consistant if as $k \to \infty$, (in Brewer's framework),

$$E(\hat{T}-T)^2 \to 0.$$

It is of interest to differentiate between the ADU strategies, using the same set of π-values, by studying for example, their second-order efficiency (in some sense).

5. Optimality of a linear predictor under multiple regression models

Pfefferman (1984), Tam (1986, 1988), Brewer, Tam and Hanif (1988), among others, addressed the question when a linear predictor becomes BLUP under $\eta(\boldsymbol{X}, \boldsymbol{V})$. It has been proved (Tam, 1986),

Theorem 5.1. *The n.s. conditions for $\boldsymbol{h}'\ \boldsymbol{Y}_s$ when $\boldsymbol{h} = (h_{1s}, \ldots, h_{ns})$ to be BLUP of Y under $\eta(\boldsymbol{X}, \boldsymbol{V})$ are : for all $s : p(s) > 0$,*

(i) $\boldsymbol{h}'\, \boldsymbol{X}_s = \boldsymbol{X}$

(ii) $\boldsymbol{V}_s\, \boldsymbol{h} - \boldsymbol{K}\, \boldsymbol{1} \in \mathcal{C}(\boldsymbol{X}_s)$.

Tam examined the conditions required for different predictors e.g. $\hat{T}^*(\boldsymbol{X}, v)$, $\hat{\hat{T}}(\boldsymbol{X}, v)$, $N\bar{Y}_s$ to be BULP under $\eta(\boldsymbol{X}^*, \boldsymbol{V})$ where $\boldsymbol{X}^*$ may lack (accommodate) some (additional) explanatory variables. Bolfarine and Rodrigues (1988) examined the conditions when $\hat{T}^*(\boldsymbol{X}, \boldsymbol{V}) = \hat{\hat{T}}^*(\boldsymbol{X}, \boldsymbol{V})$.

Brewer *et al.* (1988) cosidered the conditions on $\boldsymbol{Q}$ when $\hat{T}(\boldsymbol{Q})$ [Royall's predictor using $\hat{\beta}(\boldsymbol{Q})$], $\hat{\hat{T}}(\boldsymbol{Q})$ and $\hat{T}_{GR}(\boldsymbol{Q})$ coincide under $\eta(\boldsymbol{X}, v)$ so that $\hat{T}(\boldsymbol{Q})$ and $\hat{T}(Q)$ may share the ADU and ADC properties of $\hat{\hat{T}}_{GR}(\boldsymbol{Q})$. Tam (1986) examined the s.d's for which $\hat{T}^*$, $\hat{\hat{T}}^*$, $\hat{T}^*_{GR}$ become identical.

Tam (1988) confronted a predictor to $\hat{T}_{GR}(\boldsymbol{Q})$ and investigated the conditions for a linear strategy to be asymptotically best (i.e., ADU and attaining (4.4) under $\eta(\boldsymbol{X}, v)$). He proved

Theorem 5.2. *The sufficient condition for $(\boldsymbol{h}'\ Y_s, p)$ to be best under $\eta(\boldsymbol{X}, v)$ are : for all $s : p(s) > 0$,*

(a) (i) of Theorem 5.1 holds

(b) $\boldsymbol{Q}_s^{-1}\,[\boldsymbol{h} - h\boldsymbol{V}_s^{-1/2}\,\boldsymbol{1}_s] \in \mathcal{C}(\boldsymbol{X}_s)$

where $k = \frac{1}{n} \Sigma \sqrt{v_i}$, $\boldsymbol{Q}$ *is any p.d. diagonal matrix.*

A predictor satisfying (b) has been termed weakly robust (or robust under covariance-matrix mis-specification). Relevance of these conditions for robust prediction under $\eta(\boldsymbol{X},v)$ have been considered.

6. Sampling designs making predictors robust and near optimal

The approach in this section is to find s.d.s within a class of competing designs which endow robustness and near-optimality to some simple predictors—optimality being in the sense of attaining GJLB. Noting that under model $\eta_\beta(a) : \mathcal{E}(Y_i) = a_i+\beta x_i$, $\mathcal{V}(Y_i) = \sigma^2 f(x_i)$, $\mathcal{C}(Y_i, Y_{i'}) = 0\ (i \neq 1')$, $(a_i, \beta, f(x_i)$ known quantities), the generalised difference predictor

$$e(\boldsymbol{a}, \beta) = \sum_s \frac{Y_i-a_i-\beta x_i}{u_i}+a+\beta X, \qquad \ldots \quad (6.1)$$

$a = \Sigma\, a_i$, coupled with a $p \in \rho_\pi : \{p : \pi_i(p) \propto f(x_i)\}$ is optimum in the class of p-unbiased predictors, Godambe and Thompson (1977), with a view to choosing a suitable p in μ_π, considered the absolute difference in magnitude of the average variances of a strategy under two models as a measure of robustness (or sensitiity) of the strategy to model-changes and showed that a $p \in \mu_\pi$:

$$\sum_{i<j}\sum (\pi_i\pi_j-\pi_{ij}) \left(\frac{D_i}{\pi_i}-\frac{D_j}{\pi_j}\right)^2,$$

$$D_i = \mathcal{E}'(Y_i)-a_i-\beta x_i,$$

is minimum, is most robust wrt alternatives $\eta'_\beta : \mathcal{E}'(Y_i) \neq a_i+\beta x_i$, variance, covariance remaining the same. Such a design can be approached by an appropriate stratification. They also extended the similar study to ratio predictor.

Godambe (1982) considered the model $C : \eta_\beta(\boldsymbol{a})$, $\boldsymbol{a} \in A\ (\supset \boldsymbol{0})$, an N-dim. interval, $\beta \in B\ (\ni 0)$, an interval in $\boldsymbol{R}_1$. Noting that $(\hat{T}_{HT}, \pi p \sqrt{f(x)})$ is optimal wrt $\eta_0(\boldsymbol{0})$, he wanted to extend the optimality of $\hat{T}_{TH}$ to the entire class C by choosing a $\pi p\sqrt{f(x)}$—s.d. p_0 for the prior $\eta_\beta(\boldsymbol{a})$ such that

$$E_{p_0}[e\,(\boldsymbol{a}, \beta)-e\,(\boldsymbol{0}, 0)]^2 \qquad .$$

is very small.

Such a choice endows both 'criterion—robustness' and 'efficiency robustness' to $(p_0, \hat{T}_{HT})$ wrt alternative $\eta_\beta(\boldsymbol{a})$. p_0 can often be attained through appropriate stratifications. Iachan and Francisco (1983) carried out some empirical investigations of relative efficiencies of some of these strategies.

7. Optimal prediction under balanced samples

Mukhopadhyay and Bhattacharyya (1994) proved the following results. We denote $V(x) = \sum_k^L \delta_j \sigma_j^2 x^j$ as $V_{k,L}(x)$, $\eta(\delta_0, \ldots, \delta_J ; V_{k,J})$ as $\eta_{k,J}$, $\mathcal{E}$, $\mathcal{V}$ with proper suffix expectation and variance under the corresponding model. It follows from Theorem 2.2 that, $\hat{T}_1^*$ at $S_b(J)$, $\hat{T}_2^*$ at $S_0(J)$ are each BLUP of Y. Now

$$\mathcal{V}_{1,J-1}(\hat{T}_g^*-Y) = \sum_1^{J-1} \delta_i \sigma_i^2 x_{\bar{s}} \left[\frac{x_{\bar{s}}}{(\sum_s x_k^{2-g})^2} \sum_s x_k^{j+2-2g} + \frac{\sum_s x_k^j}{x_{\bar{s}}} \right], g \in [0, 2].$$

Thus

$$\mathcal{V}_{1,J-1}(\hat{T}_1^*-Y \mid S_b(J)) = \frac{N(N-n)}{n} \sum_1^{J-1} \delta_i \sigma_i^2 \bar{X}^{(j)} = M_1(1, J-1) \text{ (say)}$$

$$\mathcal{V}_{1,J-1}(\hat{T}_2^*-Y \mid S_0(J)) = \frac{N(N-n)}{n} \bar{x}_s \sum_1^{J-1} \delta_i \sigma_i^2 \bar{X}^{(j-1)} = M_2(1, J-1) \text{ (say)}.$$

Since under $S_0(J)$, $\bar{x}_s \leqslant \bar{X}$ [follows from (2.2) for $v(x) = x^2, j = 0$],

$$M_2 \leqslant \frac{N(N-n)}{n} \bar{X} \sum_1^{J-1} \delta_i \sigma_i^2 \bar{X}^{(j-1)}.$$

Thus

$$M_1 - M_2 \geqslant \frac{N(N-n)}{n} \sum_1^{J-1} \delta_i \sigma_i^2 [\bar{X}^{(j)} - \bar{X}\bar{X}^{(j-1)}] \geqslant 0. \quad \ldots \quad (7.1)$$

Again, under $\eta_{0,J+1}$, for HT-predictor (HTP) $\hat{T}_{HT} = \sum_s \frac{Y_i}{\pi_i}$,

$$\mathcal{V}(\hat{T}_{HT} - Y \mid S_\pi(J))$$
$$= N^2 \sum_0^{J-1} \delta_i \sigma_i^2 \left[\frac{\bar{X}\bar{X}^{(j-1)}}{n} - \frac{\bar{X}^{(j)}}{N} \right]$$
$$= M_H(0, j+1).$$

$$M_H(1, j+1) - M_2(1, j+1) = N \sum_1^{J-1} \delta_i \sigma_i^2 [\bar{x}_s \bar{X}^{(j-1)} - \bar{X}^{(j)}] \quad \ldots \quad (7.2)$$

where s refers to $S_0(J)$.

Now at $S_0(J)$,

$$N \bar{X}^{(j)} = X \bar{x}_s^{(j-1)} + \sum_s x_k^j - \bar{x}_s \sum_s x_k^{j-1}$$

Hence (7.2) simplifies to

$$N \sum_{1}^{J-1} \delta_i \sigma_i^2 \left[\frac{1}{n} \; x_{\bar{s}}(\bar{x}_s \sum_s x^{j-2} - \sum_s x_k^{j-1}) - (\sum_s x_k^j - \bar{x}_s \sum_s x_k^{j-1}) \right]$$

$$= -N \sum_{1}^{J-1} \delta_i \sigma_i^2 [x_{\bar{s}} \operatorname{cov}(x_k, x_k^{j-2}) + n \operatorname{cov}(x_k, x_k^{j-1})[\leqslant 0 \text{ if } \delta_1 = 0 \; \ldots \quad (7.3)$$

where cov denotes sample covariance. Combining (7.1) and (7.3) we have

Theorem 7.1. *Under $\eta_{2,J-1}$, $\hat{T}_{HT}$ at π-balance $S_\pi(J)$ is a better strategy than $\hat{T}_2^*$ at over-balance $S_0(J)$ which in turn is a better strategy than $\hat{T}_1^*$ at balance $S_b(J)$.*

It has been noted in section 2 that for a sample drawn by a πps-design the π-balanced properties are satisfied on expectation. This coupled with the results of Theorem 7.1 suggests that on an average, a HTP together with a πps-design will provide a better strategy than the model-dependent optimal strategies $\hat{T}_2^*$ based on an over-balanced sample and the ratio predictor $\hat{T}_1^*$ on balanced sample under the general class of polynomial regression models $\eta \left(\delta_0, \ldots, \delta_J \,;\, \sum_{1}^{J-1} \delta_i \sigma_i^2 x^i \right)$. This result provides some justification for use of a suitable design-based strategy in preference to model-dependent strategies and the role of design-based randomisation in survey sampling.

References

Arnold, S. F. (0979). Linear models with exchangeably distributed errors. *Jour. Am. Stat. Assoc.*, **74**, 194-199.

Bolfarine, H. and Rodrigues, J. (1988). On the simple projection predictors in finite popularions. *Aust. J. Statist.*, **30(3)**, 338-341.

Brewer, K. R. W. (1963). Ratio estimation and finite population : Some results deducible from the assumption of an underlying stochastic process. *Aust. J. Statist*, **5**, 93-105.

——— (1979). A class of robust sampling designs for large scale surveys. *Jour. Am. Stat. Assoc.*, **74**, 911-915.

Brewer, K. R. W., Samiuddin, M. and Asad, H. (1989). Small sample performance of five unequal probability sampling estimators (manuscript seen by courtsey of the authors).

Brewer, K. R. W., Hanif, M. and Tam (1988). How nearly can model-based prediction and design-based estimation be reconciled ? *Jour. Am. Stat. Assoc.*, **83**, 128-132.

Cassel, C. M., Sarndal, C. E. and Wretman, J. H. (1976). Some results on generalised difference estimation and generalised regression estimation for finite populations. *Biometrika*, **63**, 615-620.

Cochran, W. G. (1977). *Sampling Techniques*, 3rd edn. New York, Wiley.

Cumberland, W. G. and Royall, R. M. (1981). Prediction models and unequal probability sampling. *J. Roy. Stat. Soc.* B43, **76**, 353-367.

FULLER, W. A. and ISAKI, C. T. (1981). S urvey design under superpopulation models in *Current Topics in Survey Sampling* ed. D. Kreswski, J. N. K. Rao and R. Plateck, N. Y., Academic Press, 199-226.

GODAMBE, V. P. (1982). Estimation in survey sampling : robustness and optimality. *Jour. Am. Stat. Assoc.* **77**, 393-406.

GODAMBE, V. P. and JOSHI, V. M. (1965). Admissibility and Bayes estimation in sampling finite populations. *Ann. Math. Stat.*, **36**, 1707-22.

GODAMBE, V. P. and THOMPSON, M. E. (1977). Robust near optimal estimation in survey practice. *Bull. Int. Statist. Inst.*, **47** (3), 129-146.

HÁJEK, J. (1971). Comments on essay on the logical foundation of survey sampling, part one, by D. Basu in *Foundations of Statistical Inference* ed. Godambe, V. P. and Sprott, P. A. Holt, Rinehart and Winston, Toronto.

HERSON, J. (1976). An investigation of relative effeciency of least square prediction to conventional probability sampling plan. *Jour. Am. Stat. Assoc.* **71**, 700-703.

IACHAN, R. (1985). Robust design for ratio and regression estimation. *Jour. Stat. Pl. Inference*, **11**, 149-161.

IACHAN, R. and FRANCISCO, C. (1983). Model based sampling strategies. Unpublished manuscript.

ISAKI, C. T. and FULLER, W. A. (1982). Survey design under the regression superpopulation model. *Jour. Am. Stat. Assoc.* **73**, 89-96.

KOTT, P. S. (1986a). When a mean of ratios is the best linear unbiased estimator under a model. *Ann. Stat.* **40**(3), 202-204.

——— (1986b). Some asymptotic results for the systematic and stratified sampling of a finite population. *Biometrika*, **73**, 485-491.

LI, K. C. (1983). Consistent asymptotically efficient sampling strategies. Unpublished notes.

MONTANARI, C. E. (1987). Post-sampling efficient QR-prediction in large sample survey. *Int. Stat. Rev.* **55**(2), 191-202.

MUKHOPADHYAY, P. (1977). Robust estimation of finite population total under certain linear regression models. *Sankhyā* C, **39**, 71-87.

——— (1985a). Estimation under linear regression models. *Metrika*, **33**, 129-134.

——— (1985b). Asymptotic properties of some BLU-estimators. *Jour. Orissa Math. Soc.* **4**, 105-113.

——— (1991). Optimal estimation of a finite population total under exchangeable general linear models, *Jour. Ind. Stat. Assoc*,. **29**, 79-88.

MUKHOPADHYAY, P. and BHATTACHARYYA, S. (1994). Predicting a population total under balanced samples, to appear in *Jour. St. Pl. Inf.*

PEREIRE, C. A. B. and RODRIGUES, J. (1983). Robust linear prediction in finite population. *Int. Statist. Rev.* **51**, 293-300.

PFEFFERMAN, D. (1984). A note on large sample properties of balanced sample. *J. Roy. Stat. Soc. B***46**, 38-41.

RAO, T. J. (1984). On Brewer's class of sampling design for large scale surveys. *Metrika*, **31**, 319-322.

ROBINSON, P. M. and SARNDAL, C. E. (1983). Asymptotic properties of the generalised regression estimator in probability sampling. *Sankhyā* B, **45**, 240-248.

ROBINSON, P. M. and TSUI, K. W. (1979). On Brewer's asymptotic analysis in robust sampling designs for large scale surveys. *Tech. Rep.* No. 79-43, University of British Columbia.

——— (1981). Optimal asymptotically design unbiased estimation of a population total. *Teeh. Rep.* 634, Deptt. of Statistics, Univ. of Wisconsion, Madison.

RODRIGUES, J., BOLFARINE, H. and ROGATK, A. (1985). A general theory of prediction in finite populations. *Int. Stat. Rev.*, **53**(2), 239-54.

ROYALL, R. M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, **57**, 377-387.

ROYALL, R. M. (1975). The likelihood principle in finite population sampling theory. **40*th.*** *Session Int. Statist. Inst.*

——— (1976). The linear least-square prediction approach to two stage sampling. *Jour. Am. Stat. Assoc.* **71**, 657-664.

ROYALL, R. M. and HERSON, J. (1973). Robust estimation in finite populations I, II, *Jour. Am. Stat. Assc.* **68**, 880-889, 890-893.

ROYALL R. M. and PFEFFERMAN, D. (1982). Balanced samples and robust Bayesian inference in finite population sampling. *Biometrika*, **69**, 401-409.

SARNDAL, C. E. (1980a). Two model-based inference arguments in survey sampling. *Aust. J. Statist.* **22**, 341-358.

——— (1980b). A two-way classification of regression estimation strategies in probability sampling. *Canadian J. of Stat.* **8**(2), 165-177.

——— (1980c). On π-inverse weighting versus best linear weighting in probability sampling. *Biometrika*, **67**(3), 639-650.

SCOTT, A. J., BREWER, K. R. W. and HO, E. W. H. (1978). Finite populations sampling and robust estimation. *Jour. Am. Stat.* **73**, 359-361.

SKINNER, C. J. (1983). Multivariate prediction from selected samples. *Biometrika*, **70**, 289-92.

TALLIS, G. M. (1978). Note on robust estimation in finite populations. *Sankhyā* C, **40**, 136-138.

TAM, S. M. (1986). Characterisation of best model-based predictors in survey sampling. *Biometrika* **73**, 232-235.

——— (1988a). Asymptotic design-unbiased predictors in survey sampling, *Biometrika*, **75**, 175-177.

——— (1988b). Some results on robust estimation in finite population sampling. *Jour. Am. Stat. Assoc.* **83**, 262-8.

WRIGHT, R. L. (1983). Sampling design with multivariate auxiliary information. *Jour. Am. Stat. Assoc.* **78**, 879-884.

COMPUTER SCIENCE UNIT
INDIAN STATISTICAL INSTITUTE
203 B. T. ROAD
CALCUTTA 700035
INDIA.