Sankhyā : The Indian Journal of Statistics
1990, Volume **52**, Series B, Pt. 3, pp. 343-346.

ON ASYMPTOTIC PROPERTIES OF A GENERALISED PREDICTOR OF FINITE POPULATION VARIANCE

By P. MUKHOPADHYAY

Indian Statistical Institute

SUMMARY. A predictor of a finite population variance under probability sampling suggested by a multiple regression model is shown to be asymptotically design unbiased and consistent.

1. Introduction

We consider estimating a finite population variance through probability sampling. Let U denote a finite population of N identifiable units labelled $1, 2, \ldots, N$ and y the character of interest taking value y_i on unit i, $i = 1, \ldots, N$. Its variance is

$$V(\boldsymbol{y}) = a_1 \sum_{1}^{N} y_i^2 - a_2 \sum_{i \neq i'=1}^{N}\sum y_i y_{i'} \qquad \ldots \quad (1.1)$$

where $a_1 = \frac{1}{N}\left(1-\frac{1}{N}\right)$, $a_2 = \frac{1}{N^2}$ and $\boldsymbol{y} = (y_1, \ldots, y_N)$. Let a sample s be selected from U following a design p, having inclusion–probabilities $\pi_i = \sum_{s \ni i} p(s)$, $\pi_{ii'} = \sum_{s \ni i, i} p(s)$, etc. Let I_i, $I_{ii'}$ denote indicator random variables with $I_i = 1(0)$ according as unit $i \in (\notin) s$ and $I_{ii'} = 1(0)$ according as the pair $(i, i') \in (\notin) s$. Suppose auxiliary variable x_j with x_{ij} its value on unit i is available. Also assume that y_i is the realised value of a random variable Y_i, $i = I, \ldots, N$. We propose a predictor of $V(\boldsymbol{Y})$ where $\boldsymbol{Y} = (Y_1, \ldots, Y_N)$ as

$$v_G(\boldsymbol{Y}) = a_1 \sum_{i=1}^{N} \frac{I_i Y_i^2}{\pi_i} - a_2 \sum_{i \neq i'=1}^{N}\sum \frac{I_{ii'} Y_i Y_{i'}}{\pi_{ii'}}$$

$$+ \sum_{j=1}^{k} \hat{\beta}_j \left\{ a_1 \sum_{i=1}^{N} \left(\frac{I_i}{\pi_i} - 1 \right) x_{ij}^2 - a_2 \sum_{i \neq i'=1}^{N}\sum \left(\frac{I_{ii'}}{\pi_{ii'}} - 1 \right) x_{ij} x_{i'j} \right\}, \ldots \quad (1.2)$$

here $\hat{\beta}_j$ is a function of $\boldsymbol{I}$, $\boldsymbol{Y}$ and $\boldsymbol{X}$, $\boldsymbol{I} = (I_1, \ldots, I_N)'$, $\boldsymbol{X} = ((x_{ij}))$ an $N \times k$ matrix such that $\hat{\beta}_j$ when suitably assigned is computable given the data stated above. The multiple-regression model-based form (1.2) is suggested following Särndal (1980).

AMS (1980) *subject classification :* 62D05.

Key words and phrases : Multiple regression model, Generalised predictor, Asymptotic design unbiasednes, Asymptotic design consistency.

Following Isaki and Fuller (1982) and Robinson and Särndal (1983) we show that $v_G(Y)$ is asymptotically design unbiased and consistent for $V(Y)$ under conditions which do not require any modelling.

2. Asymptotic design unbiasedness and consistency of the generalised predictor

Following Robinson and Särndal (1983) we define a sequence of populations U_t of increasing sizes $N_1 < N_2 < N_3 < \ldots$ such that $U_1 \subset U_2 \subset U_3 \ldots$ Let $\{s_t\}$ denote a sequence of samples s_t of effective size n_t drawn from U_t using sampling design p_t, $t = 1, 2, 3, \ldots$ with $n_1 < n_2 < n_3 < \ldots$. Let π_{it}, $\pi_{ii't}$ etc. denote inclusion-probabilities for p_t. Let also I_{it} and $I_{ii't}$ denote corresponding indicator variables. Then we have a sequence of population values $\{y^t, X^t\}$ where $y^t = (y_1, \ldots, y_{N_t})$, $X^t = ((x_{ij}))$ is an $N_t \times k$ matrix, a sequence of popuation parameters $\{V_t(y^t)\}$ and a sequence of predictors $\{v_{G_t}(Y^t)\}$ where

$$v_{G_t}(Y^t) = a_{1t} \sum_{i=1}^{N_t} \frac{I_{it} Y_i^2}{\pi_{it}} - a_{2t} \sum_{i \neq i'=1}^{N_t}\sum \frac{I_{ii't} Y_i Y_{i'}}{\pi_{ii't}}$$

$$+ \sum_{j=1}^{k} \hat{\beta}_{jt} \left\{ a_1 \sum_{j=1}^{N_t} \left(\frac{I_{ij}}{\pi_{it}} - 1 \right) x_{ij}^2 \right.$$

$$\left. - a_{2t} \sum_{i \neq i'=1}^{N_t}\sum \left(\frac{I_{ii't}}{\pi_{ii't}} - 1 \right) x_{ij} x_{i'j} \right\}, \qquad \ldots \quad (2.1)$$

$a_{1t} = \frac{1}{N_t}\left(1 - \frac{1}{N_t}\right)$, $a_{2t} = \frac{1}{N_t^2}$, $\hat{\beta}_{jt}$ is a function of I_t, Y^t and X^t with $I_t = (I_{1t}, \ldots, I_{N_t t})'$ and $Y^t = (Y_1, \ldots, Y_{N_t})$.

For the asymptotic analysis let $N_t \to \infty$ as $t \to \infty$. Let ξ be the probability distribution of the infinite dimensional random vector $(Y_1, Y_2, \ldots)$.

Definition 1. $\{v_{Gt}\}$ is asymptotically design unbiased (ADU) if

$$\lim_{t \to \infty} E\{(v_{Gt} \mid Y^t) - V(Y^t)\} = 0$$

with ξ—probability one.

Definition 2. $\{v_{Gt}\}$ is asymptotically design consistent (ADC) for V_t if given any $\epsilon > 0$,

$$\lim_{t \to \infty} P\{|v_{Gt} - V_t| > \epsilon \mid Y^t\} = 0$$

with ξ—probability one.

Here E denotes design expectation. By Markov's inequality if v_{Gt} is *ADU* it must be *ADC*.

Theorem : *Under assumptions* (*a.* 1)—(*a.* 9) *below*, $\boldsymbol{v}$ *is ADU and ADC. The assumptions are* :

(*a.* 1) $\overline{\lim_{t \to \infty}} \frac{1}{N_t} \sum_{i=1}^{N_t} Y_i^4 < \infty$ *with* ξ*-probability one.*

(*a.* 2) $\underline{\lim_{t \to \infty}} \phi_1(t) = \infty$ *where* $\phi_1(t) = N_t \min_{1 \leqslant i \leqslant N_t} \pi_{it}$.

(*a.* 3) $\lim_{t \to \infty} \psi_1(t) = 0$ *where* $\psi_1(t) = \max_{1 < i \neq i' \leqslant N_t} \left| \frac{\pi_{ii't}}{\pi_{it}\pi_{i't}} - 1 \right|$

(*a.* 4) $\underline{\lim_{t \to \infty}} \phi_2(t) = \infty$ *where* $\phi_2(t) = N_t^2 \min_{1 \leqslant i \neq i' \leqslant N_t} \pi_{ii't}$

(*a.* 5) $\lim_{t \to \infty} \psi_2(t) = 0$ *where* $\psi_2(t) = \frac{1}{N_t} \max_{1 < i \neq i' \neq i'' \leqslant N_t} \left| \frac{\pi_{ii'i''t}}{\pi_{ii't}\pi_{ii''t}} - 1 \right|$

(*a.* 6) $\lim_{t \to \alpha} \psi_3(t) = 0$ *where* $\psi_3(t) = \max_{1 \leqslant i \neq i' \neq i'' \neq i''' \leqslant N_t} \left| \frac{\pi_{ii'i''i'''t}}{\pi_{ii't}\pi_{i''i'''t}} - 1 \right|$

(*a.* 7) $\lim_{t \to \alpha} \psi_4(t) = 0$ *where* $\psi_4(t) = \max_{1 \leqslant i \neq i' \neq i'' \leqslant N_t} \left| \frac{\pi_{ii'i''t}}{\pi_{it}\pi_{i'i''t}} - 1 \right|$

(*a.* 8) $\overline{\lim_{t \to \alpha}} \frac{1}{N_t} \sum_{i=1}^{N_t} x_{ij}^4 < \infty$ *for* $j = 1, 2, \ldots, k$.

(*a.* 9) $\overline{\lim_{t \to \alpha}} E \left(\sum_{j=1}^{k} \hat{\beta}_{jt}^2 \right) < \alpha$ *with* ξ*-probability one.*

Proof : We have

$$v_{Gt} - V_t = C_t(\boldsymbol{y}) + \sum_{j=1}^{k} \hat{\beta}_{jt} C_t(x_j) \qquad \ldots \quad (2.2)$$

where

$$C_t(\boldsymbol{y}) = a_{1t} \sum_{i=1}^{N_t} Y_i^2 \left(\frac{I_{it}}{\pi_{it}} - 1 \right) - a_{2t} \sum_{i \neq i'}^{N_t} \sum_{=1} Y_i Y_{i'} \left(\frac{I_{ii't}}{\pi_{ii't}} - 1 \right)$$

and $C_t(x_j)$ is defined similarly. Hence

$$E\{|v_{Gt} - V_t| \, | \boldsymbol{Y}^t\} \leqslant \sqrt{E(C_t^2(\boldsymbol{y}) | y^t)} + \sqrt{E\left(\sum_{j=1}^{k} \hat{\beta}_{jt}^2 \, | \boldsymbol{Y}^t \right) E\left(\sum_{j=1}^{k} C_t(x_j)^2 \right)} \qquad \ldots \quad (2.3)$$

Now

$$E(C^2_{t}(y)\,|\,Y^t) = a^2_{1t}\Big[\sum_{i=1}^{N_t} Y_i^4\left(\frac{1}{\pi_{it}}-1\right) + \sum_{i\neq i'=1}^{N_t}\sum Y_i^2 Y_{i'}^2 .$$

$$\left(\frac{\pi_{ii't}}{\pi_{it}\pi_{i't}}-1\right)\Big] + a^2_{2t}\Big[2\sum_{i\neq i'=1}^{N_t}\sum Y_i^2 Y_{i'}^2\left(\frac{1}{\pi_{ii't}}-1\right)$$

$$+4\sum_{i\neq i'\neq i''=1}^{N_t}\sum\sum Y_i^2 Y_{i'} Y_{i''}\left(\frac{\pi_{ii'i''t}}{\pi_{ii't}\,\pi_{ii''t}}-1\right)$$

$$+\sum_{i\neq i'\neq i''\neq i'''=1}^{N_t}\sum\sum\sum Y_i Y_{i'} Y_{i''} Y_{i'''}\left(\frac{\pi_{ii'i''i'''t}}{\pi_{ii't}\,\pi_{i''i'''t}}-1\right)\Big]$$

$$-2\,a_{1t}a_{2t}\Big[2\sum_{i\neq i'=1}^{N_t}\sum Y_i^3 Y_{i'}\left(\frac{1}{\pi_{it}}-1\right)$$

$$+\sum_{i\neq i'\neq i''=1}^{N_t}\sum\sum Y_i^2 Y_{i'} Y_{i''}\left(\frac{\pi_{ii'i''t}}{\pi_{it}\pi_{i'i''t}}-1\right)\Big] \qquad \ldots \quad (2.4)$$

The first term in (2.4) is dominated by

$$\frac{1}{N_t}\sum_{i=1}^{N_t}\frac{Y_i^4}{\phi_1(t)}$$

and $\to 0$ as $t \to \infty$ with ξ-probability one under assumptions (a. 1) and (a.2). The subsequent terms also tend to 0 with ξ-probability one as $t \to \infty$ under (a. 1)—(a. 7). Hence $E(C^2{}_t(y)\,|\,Y^t) \to 0$ with ξ-probability one as $t \to \infty$. Similarly under (a. 2)—(a. 8), $E\sum_{j=1}^{k} C^2{}_t(x_j) \to 0$ as $t \to \infty$. These coupled with the assumption (a. 9) prove that v_{Gt} is *ADU* and *ADC*.

Note : The assumptions (a. 2), (a. 4), (a. 5) imply $n_t \to \infty$ as $t \to \infty$. All the assumptions (a. 2)—(a. 7) are satisfied for simple random sampling.

Acknowledgement. Thanks are due to a referee for some suggestions for improvement in presentation.

REFERENCES

ISAKI C. T. and FULLER, W. A. (1982) : Survey design under the regression super-population model. *Jour. Am. Stat. Assoc.* **77**, 89-96.

MUKHOPADHYAY, P. (1986) : Asymptotic properties of a generalised predictor of finite population variance under probability sampling. *Ind. Stat. Instt. Tech. Rep.* No. ASC/86/19.

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

SÄRNDAL, C. E. (1980) : On π-inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrika*, **67**, 639-650.

Paper received : *October*, 1987.

Revised : *November*, 1988.