STATISTICS IN TRANSITION-new series, December 2009
Vol. 10, No. 3, pp. 415—435

FULL INFORMATION EFFICIENT ESTIMATOR OF FINITE POPULATION VARIANCE

Manoj Kumar Srivastava[1], Namita Srivastava[2] , Housila P. Singh[3]

ABSTRACT

Second order or quadratic and finite population parametric functions may be expressed as total of variable-values on pairs of units in a derived population. Recently, Sitter and Wu (2002) utilized this approach for estimating variance under model calibration. In this paper an efficient design based full-information estimator of finite population variance has been suggested. The exact expression of its variance and its relative efficiency has also been derived. Finally, the proposed estimator has been shown to be superior to its competitors in an empirical investigation.

Key words: design based estimation; variance estimation; Rao-Blackwellization in survey sampling; estimation of polynomial finite population function.

1. Introduction

In finite population theory the problem of estimating quadratic or higher order finite population functions is an extensively explored area of research. Efficient estimators for finite population variances, covariance between two response variables or variance of linear estimators are highly desirable. Efficient estimation of these quadratic functions have been prohibitive (Sitter and Wu 2002), because of complex expressions of variances of these estimators. Liu (1974 a) suggested several design based estimators of finite population variance following which Chaudhuri (1978) noted that many of Liu's estimators sometimes take negative values. He then suggested alternatively few non-negative estimators and discussed their statistical properties. Later, Liu and Thompson (1983) showed nonexistence of the best unbiased estimator of population variance and asserted

[1] Associate Professor, Department of Statistics, Institute of Social Sciences, Dr. BRA University, Agra-282002.U.P. India, email : mks_iss@yahoo.co.in.
[2] Associate Professor, Department of Statistics, St. John's College, Agra-282002, U.P. India. email: drnamita.sjc@gmail.com.
[3] Professor, School of Studies in Statistics, Vikram University, Ujjain-456010, M.P. India, email: hpsujn@ rediffmail.com.

that the admissibility holds under certain restrictions. Swain and Mishra (1994) suggested more efficient estimator of population variance as compared to Liu's and Chaudhary's estimators and tried it on various natural and artificial populations. This estimator still suffers from a drawback that it takes negative values occasionally. Mukhopadhyay (1978) suggested model based predictors of population variance following the Royall's prediction approach. Mukhopadhyay and Bhattacharya (1989) suggested optimal estimators of population variance under regression superpopulation models. Sitter and Wu (2002) for the first time suggested very efficient model calibration estimators which were asymptotically design unbiased under linear and non-linear models. They suggested pseudoempirical likelihood estimators also which are free from taking negative values.

Bayesian estimation of finite population variance was first taken up by Ericson (1969) in his pioneering work on Bayesian estimation in survey sampling although the issue of estimation of population variance was briefly discussed. Subsequent references are Liu (1974 b), Chaudhuri (1978), Zacks and Soloman (1981), Ghosh and Meeden (1983, 1984),

Vardeman and Meeden (1983), Ghosh and Lahiri (1987), Lahiri and Tiwari (1991), Datta and Tiwari (1991), Datta and Ghosh (1993) among others.

Hanurav (1966) was first who attempted simplification of the problem of estimation of population variance by expressing it as a total of a function over a new population, whose elements are ordered pairs of units of the original population. Using this approach, Liu and Thompson (1983) showed nonexistence of best unbiased estimators of population variance and gave some admissibility results. More recently, Sitter and Wu (2002) named the population of pairs as "synthetic population" and under the Hanurav's approach they discussed model calibration and pseudoempirical likelihood estimators which were model assisted and asymptotically design unbiased under a general sampling design. We have suggested an efficient estimator named as full-information estimator of population variance for equal and unequal probability sampling design under the above set up by drawing samples from "synthetic population". We have used this estimator on a population of 124 countries (CO124, Sarn-dal, Swensson, Wretmen 1992) for estimating the variance of imports among 124 countries, and have shown that the proposed estimator is more efficient as compared to its competitors in terms of relative efficiency (RE).

1.1.Notations

We denote population by $U = \{U_1, U_2, \cdots, U_N\}$; and the population of ordered pairs by $U' = \{\mathbf{U}'_k : \mathbf{U}'_k = (U_{i1}, U_{i2}), U_{i1}, U_{i2}$ are units taken from U so that $i_1 < i_2\}$. The value of the study variable y on U_i be denoted by y_{i1} and on $\mathbf{U}'_\mathbf{k}$ be denoted by (y_{i1}, y_{i2}). Define, for a real symmetric function h on (y_{i1}, y_{i2}), $t_k = h(y_{i1}, y_{i2})$ for some k in U', $\overline{T}_{U'} = \frac{1}{N'}\sum_{k=1}^{N'} t_k$, $N' = \binom{N}{2}$. $\overline{T}_{U'}$ for different h functions:

1. For $t_k(y) = \frac{1}{2}\left(y_i - y_j\right)^2, k = (i, j)$, we get $\overline{T}_{U'} = \frac{1}{N-1}\sum_U (y_i - \overline{y})^2 = S_y^2$: finite population variance.
2. For $t_k(y, z) = \frac{1}{2}\left(y_i - y_j\right)\left(z_i - z_j\right) = C_{yz}$, we get $\overline{T}_{U'} = \frac{1}{N-1}\sum_U (y_i - \overline{y})(z_i - \overline{z})$: population covariance.
3. For $t_k(y) = \frac{N(N-1)}{2}\left(\pi_i\pi_j - \pi_{ij}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$, we get $\overline{T}_{U'} = \sum_{i<j\in U} (\pi_i\pi_j - \pi_{ij})\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 = V_{YG}$, which is Yates and Grundy-type variance of Horvitz Thompson (HT) estimator of a population total, $\widehat{Y}_{HT} = \sum_s \frac{y_i}{\pi_i}$
4. For $t_k = \frac{|y_i - y_j|}{2\overline{y}}$, gives $\overline{T}_{U'} = \sum_{i<j\in U} \frac{|y_i - y_j|}{N(N-1)\overline{y}} = G$, which is population Gini-coefficient.

Note that in (1) the finite population variance S_y^2 has been expressed as mean of t_k values over the population U'.

2. Estimation under unequal probability design

Associated with U_i let x_i be the size measure so that $y_i \propto x_i$; $P_i = x_i / x$, $x = \sum_U x_k$. Note for the function h in (1) is such that $t_k(\mathbf{y}) \propto t_k(\mathbf{x})$; $t_k(\mathbf{x}) = \frac{1}{2}\left(x_i - x_j\right)^2$ whenever $y_i \propto x_i$. Let the size measures on U' be $t_k(\mathbf{x})$, so that $P'(U'_k) = \frac{t_k(\mathbf{x})}{\sum_{U'} t_k(\mathbf{x})}$. We consider here a probability proportional to size design with replacement with sample size *n(s) = n*. Let the corresponding sample space be denoted by S'. Denote this design by $D'(U', S', P')$. Based on ppswr sample s_n of size n and the corresponding data $d = \left\{\left(\mathbf{U}'_k, t_k\right) : \mathbf{U}'_k \in s_n\right\}$, the unbiased estimator of $\overline{T}_{U'}\left(S_y^2\right)$ is given by

$$\hat{\bar{T}}_{s_n} = \frac{1}{n}\sum_{s_n} \frac{t_k}{N'P'_k} \tag{2.1}$$

The variance of $\hat{\bar{T}}_{s_n}$ is given by

$$V\left(\hat{\bar{T}}_{s_n}\right) = \frac{1}{n}\sum_{U'} P'_k P'_l \left(\frac{t_k}{N'P'_k} - \frac{t_l}{N'P'_l}\right)^2 \tag{2.2}$$

Where $\sum_{U'}$ is summation over all different pairs of 2-tuples in U'. Furthermore the estimator of $V(\cdot)$ is given by

$$v\left(\hat{\bar{T}}_{s_n}\right) = \frac{1}{n^2(n-1)}\sum_{s_n} \left(\frac{t_k}{N'P'_k} - \frac{t_l}{N'P'_l}\right)^2 \tag{2.3}$$

Where the summation $\sum_{s_n}$ is over all pairs of 2-tuples in the sample s_n. Let us now consider a reducing transformation on s_n that ignores order and multiplicity of 2-tuples and results into a sample $s_{n'}$ and the corresponding data be denoted by $d^* = \{(\mathbf{U}'_k, t_k) : \mathbf{U}'_k \in s_{n'}\}$. Since the sample $s_{n'}$ is a sufficient statistic (Cassel, Sarndal, and Wretman 1976), therefore, by using Rao-Blackwellization, we get an improved unbiased estimator $\hat{\bar{T}}^*_{s_{n'}} = E^{S_n | s_{n'}}\left\{\hat{\bar{T}}_{s_n} \mid D^* = d^*\right\}$ since $E^{S_{n'}}\left\{\hat{\bar{T}}^*_{s_{n'}}\right\} = E^{S_{n'}} E^{S_n | S_{n'}}\{.\} = \bar{T}_{u'}$ with variance

$$V\left(\hat{\bar{T}}^*_{s_{n'}}\right) = E^{S_{n'}}\left\{E^{S_n | S_{n'}}\left\{\hat{\bar{T}}_{s_n} \mid D^* = d^*\right\}\right\}^2 - \bar{T}^2_{u'}$$

$$= \sum_{s_{n'}}\left[\sum_{\substack{s_n \in S' \\ \text{s.t. } s_n \text{ gives } s_{n'}}} \hat{\bar{T}}_{s_n} \cdot \frac{P'(s_n)}{P'(s_{n'})}\right]^2 P'(s_{n'}) - \bar{T}^2_{u'}$$

Where $P'(s_{n'}) = \sum_{\substack{s_n \in S' \\ \text{s.t. } s_n \text{ gives } s_{n'}}} P'(s_n)$ further $V\left\{\hat{\bar{T}}^*_{s_{n'}}\right\} \le V\left\{\hat{\bar{T}}_{s_n}\right\}$.

While utilizing information on already selected 2-tuples in a ppswr sample we get some additional information on few more 2-tuples. For example if $\mathbf{U}'_{\mathbf{i}_1} = (U_1, U_2)$ and $\mathbf{U}'_{\mathbf{i}_2} = (U_1, U_3)$ are selected in a sample, observing these 2-tuples also means observing U_1, U_2 and U_3. So we have effectively derived information on $\mathbf{U}'_{\mathbf{i}_3}$ in addition to information on $\mathbf{U}'_{\mathbf{i}_1}$ and $\mathbf{U}'_{\mathbf{i}_2}$ where, $\mathbf{U}'_{\mathbf{i}_3} = (U_2, U_3)$. We expect that this information on $\mathbf{U}'_{\mathbf{i}_3}$ could also be utilized to get a better estimator though it was not directly selected. Let the sample containing information on $\mathbf{U}'_{\mathbf{i}_1}, \mathbf{U}'_{\mathbf{i}_2}$, $\mathbf{U}'_{\mathbf{i}_3}$ be called full-information sample denoted by $s\binom{n''}{2}$ and subset of n'' distinct units be denoted by $A_{n''} = \{U_{i_1}, U_{i_2}, \cdots, U_{n''}\} = \{U_1, U_2, U_3\}$ (as in the above example). $A_{n''}$ could be derived from ppswr sample s_n or by $s_{n'}$ (with their usual notations) by a reducing function on s_n, which gives only distinct units appeared on 2-tuples in $s_n\ (s_{n'})$. The probability distribution of such a random set $A_{n''}$ has been stated in the theorem 2.1. We have $P'(s_n) = P'(\mathbf{U}'_{\mathbf{i}_1}).P'(\mathbf{U}'_{\mathbf{i}_2})....P'(\mathbf{U}'_{\mathbf{i}_n})$. Define a random set

$$A_{n''} = \mathbf{U}'_{\mathbf{i}_1} \cup \mathbf{U}'_{\mathbf{i}_2} \cup \cdots \cup \mathbf{U}'_{\mathbf{i}_n}$$

with a meaning that $A_{n''}$ is a set of distinct units U_i's appearing in one or more $\mathbf{U}'_{\mathbf{i}_j}$. Clearly, $A_{n''}$ is a random set in the language of probability theory, taking its values from U as a subset. Let us denote the number of distinct units in $A_{n''}$ by $|A_{n''}|$, this number is also an integer-valued random variable. To discuss the statistical properties of the estimator based on $A_{n''}$ we would be interested in the probability distribution of the random set $A_{n''}$. Consider a set B of M units being fixed from U as a subset, namely, $U_{l_1} \cup U_{l_2} \cup \cdots \cup U_{l_M}$; generate a set of $\binom{M}{2} = M'$ 2-tuples purely made of units in B, denote it by $U'_B, |B| = M, 1 \le M \le N$, clearly $U'_B \subset U'$. If one draws probability proportional to size samples of size n from U'_B with replacement (ppswr), the corresponding sample space be given by S'_B . Note that S'_B is a collection of all 2-tuples made up entirely of the units (labels) in the subset B so that $S'_B \subseteq S'$. The probability of any such sample $S_n^{(B)} = \{\mathbf{U}'_{\mathbf{j}_1}, \mathbf{U}'_{\mathbf{j}_2} \cdots, \mathbf{U}'_{\mathbf{j}_n}\}$, each $\mathbf{U}'_{\mathbf{j}_i}$ is some 2-tuple made up of units in B, under the design D'

$$P'\left(s_n^{(B)}\right) = P'\left(\mathbf{U}_{\mathbf{j}_1}\right).P'\left(\mathbf{U}_{\mathbf{j}_2}\right).......P'\left(\mathbf{U}_{\mathbf{j}_n}\right)$$

where, $\mathbf{U}_{\mathbf{j}_i}$'s are not necessarily distinct. The probability distribution of the set $A_{n''}$ is given by :

Theorem 2.1. Consider a probability proportional to size sample s_n of size $n(s_n) = n$ drawn from U' with replacement under the ppswr design $D'(U', S', P')$. Let the constituent 2-tuples in s_n be denoted by $\mathbf{U}'_{\mathbf{i}_1}, \mathbf{U}'_{\mathbf{i}_2}, ..., \mathbf{U}'_{\mathbf{i}_n}$. Define a set valued random variable taking its values from U,

$$A_{n''} = \mathbf{U}'_{\mathbf{i}_1} \cup \mathbf{U}'_{\mathbf{i}_2} \cup ...... \cup \mathbf{U}'_{\mathbf{i}_n}$$

as *the set of distinct units common to* $\mathbf{U}'_{\mathbf{i}_j}$ *'s. The size of* $A_{n''}$*, i.e.*| $A_{n''}$| *would range from* $\underline{k} = 2$ *to* $\overline{k} = \min\{2n, N\}$. *The probability distribution of* $A_{n''}$ *is given by*

$$P'(A_{n''})|_{\substack{A_{n''}=B,\\ |B|=\underline{k}}} = \sum_{S'_B} P'\left(s_n^{(B)}\right)$$

for every B of size $\underline{k}$. Next

$$P'(A_{n''})|_{\substack{A_{n''}=B,\\ |B|=\underline{k}+1}} = \sum_{S'_B} P'\left(s_n^{(B)}\right) - \sum_{\substack{\{A_{n''}:|A_{n''}|=\underline{k},\\ A_{n''}\subset B\}}} P'(A_{n''})$$

for every B of size $\underline{k}+1$, where, $P'(A_{n''})$ in the second term on the right hand side of the above expression comes from the previous step. Proceeding similarly, for $k = \underline{k} + j$

$$P'(A_{n''})|_{\substack{A_{n''}=B,\\ |B|=\underline{k}+j}} = \sum_{S'_B} P'\left(s_n^{(B)}\right) - \sum_{\substack{\{A_{n''}:|A_{n''}|=\underline{k}+(j-1),\\ A_{n''}\subset B\}}} P'(A_{n''})$$

for every B of size $\underline{k}+j$. Proceeding similarly up to $j = \overline{k} - \underline{k}$, one gets complete distribution of $A_{n''}$ recursively. (proof of the theorem has been given in Appendix).

It can easily be seen that $\dfrac{P'(s_n)}{P'\left[s\binom{n''}{2}\right]}$ or $\dfrac{P'(s_{n'})}{P'\left[s\binom{n''}{2}\right]}$ independent of the parameter vector, $\mathbf{t}$ = $(t_1,.......,t_{N'})'$, therefore, $s\binom{n''}{2}$ or $A_{n''}$ is a sufficient statistic.

Since, $\hat{\overline{T}}_{s_n}$ and $\hat{\overline{T}}^*_{s_{n'}}$ are both unbiased estimators of $\overline{T}_{U'}$ and $s\binom{n''}{2}$ or $A_{n''}$ is sufficient, therefore, Rao Blakwellization of both the estimators gives the same unbiased estimator $\tilde{\overline{T}}_{s\binom{n''}{2}}|_{full-\inf}$ defined as

$$\tilde{\overline{T}}_{s\binom{n''}{2}}|_{full-\inf} = \sum_{s_n \Rightarrow A_{n''}} \hat{\overline{T}}_{s_n} \cdot \frac{P'(s_n)}{P'(A_{n''})} = E^{S_n|A_{n''}}\left\{\hat{\overline{T}}_{S_n}\right\} \tag{2.4}$$

where, $\sum$ is over all such ppswr samples of size s_n in S' that result into same $A_{n''}$ given $A_{n''}$ in *U*, $S'_{A_{n''}}$, where ppswr samples s_n reduces into a given $A_{n''}$.

S_n in the above expectation is a random variable taking its values from the subspace $S'_{A_{n''}}$ This estimator is unbiased, since, $E^{A_{n''}}\left\{\tilde{\bar{T}}_{s\binom{n''}{2}} \mid full\text{-}\inf\right\} = E^{A_{n''}} E^{S_n \mid A_{n''}}\left\{\hat{\bar{T}}_{s_n}\right\} = \bar{T}_{U'}$ where $E^{A_{n''}}$ is over all the subsets of U. Further, the variance of the full-information estimator is given by

$$V\left\{\tilde{\bar{T}}_{s\binom{n''}{2}} \mid full\text{-}\inf\right\} = E^{A_{n''}}\left[E^{S_n \mid A_{n''}}\left\{\hat{\bar{T}}_{s_n}\right\}\right]^2 - \bar{T}_{U'}^2,$$

$$= \sum_{A_{n''} \subset U}\left[\sum_{\substack{s_n \in S' \text{ so that} \\ s_n \text{ gives } A_{n''} \Rightarrow s\binom{n''}{2}}} \hat{\bar{T}}_{s_n} \cdot \frac{P'(s_n)}{P'(A_{n''})}\right]^2 P'(A_{n''}) - \bar{T}_{U'}^2,$$

where, $\sum_{s_n \in S'}$ is over all such ppswr samples in S' that gives the same $A_{n''}$ which is being fixed at the previous summation $\sum_{A_{n''} \subset U}$. Here, $P'(A_{n''})$ are obtained as the probability distribution of the random sets $A_{n''}$ derived in Theorem 2.1.

Finally, we get the following relationship

$$V\left\{\tilde{\bar{T}}_{s\binom{n''}{2}} \Big| \, full\text{-}inf\right\} \le V\left\{\tilde{\bar{T}}^*_{s_{n'}}\right\} \le V\left\{\hat{\bar{T}}_{s_n}\right\} \tag{2.5}$$

3. Estimation under equal probability design

Let the design $D'(U', S', P')$ be srswr. We draw a simple random sample s_n of size n from U' with replacement (srswr) and denote the corresponding data by $d = \{(\mathrm{U}_k, t_k) : \mathrm{U}_k \in s_n\}$. Based on d we have an unbiased estimator of S_y^2

$$\hat{\bar{T}}_{s_n} = \frac{1}{n}\sum_{s_n} t_k = \frac{1}{n}\sum_{k=1}^{n'} f_k t_k \tag{3.1}$$

where f_k is the number of times the k^{th} 2-tuple is repeated in the sample s_n, and n' be the number of distinct 2-tuples in s_n. $V\left(\hat{\bar{T}}_{s_n} \mid srswr\right) = \dfrac{\binom{N}{2} - 1}{\binom{N}{2}} \cdot \dfrac{S_t^2}{n}$

where $S_t^2 = \frac{1}{\binom{N}{2}-1}\sum_{U'}\left(t_k - \bar{T}_{U'}\right)^2$. Using the result of Raj and Khamis (1958) and Des Raj (1968)-pp40, theorem 3.5, a better unbiased (Rao-Blackwellized) estimator of s_y^2 based on sample $s_{n'}$ of n' distinct 2-tuples in s_n is

$$\hat{\bar{T}}^*_{s_{n'}} = E^{S_n|s'_n}\left[\hat{\bar{T}}_{s_n}\right] = \frac{1}{n'}\sum_{s_{n'}} t_k \tag{3.2}$$

with $V\left[\hat{\bar{T}}^*_{s_{n'}}\right] = \frac{\binom{N}{2}-n'}{\binom{N}{2}n'}.S_t^2 \le V\left[\hat{\bar{T}}_{s_n}\right]$

We select n' 2-tuples from U', by srswor, consequently we get a set, $A_{n''}$, of n'' distinct units of *U*. These n'' distinct units allow us to know the information (t-values) on $\binom{n''}{2}$ distinct 2-tuples. For equal probability without replacement sampling the probability distribution of these n'' units is stated in the following corollary to theorem 2.1. In theorem 2.1 we noted that the probabilities $P'(A_{n''})$ were all different for different $A_{n''}$'s of the same size because these probabilities were depending on the units constituting $A_{n''}$ whereas for equal probability sampling the probabilities of sets $A_{n''}$ do not depend on the units constituting $A_{n''}$ rather on the size of $A_{n''}$, *i.e.*, $|A_{n''}|$.

Corollary 3.1. *Consider the design* $D'(U', S', P')$ *as simple random sampling without replacement. Consider a srswor sample* $s_{n'}$ *of size* n' *from* U' *under the design* D'*, the probability distribution of distinct units* n'' *in* $A_{n''}$ *would be given by*

$$P'(n'' = \underline{k}) = \frac{\binom{\binom{\underline{k}}{2}}{n'}\binom{N}{\underline{k}}}{\binom{\binom{N}{2}}{n'}} \tag{3.3}$$

$$P'(n'' = \underline{k}+1) = \left[\frac{\binom{\binom{\underline{k}+1}{2}}{n'}}{\binom{\binom{N}{2}}{n'}} - \frac{(\underline{k}+1)P'(n'' = \underline{k})}{\binom{N}{\underline{k}}}\right]\binom{N}{\underline{k}+1} \tag{3.4}$$

$$P'(n''=\underline{k}+j)=\left[\frac{\binom{\binom{\underline{k}+j}{2}}{n'}}{\binom{\binom{N}{2}}{n'}}-\sum_{k=\underline{k}}^{\underline{k}+j-1}\frac{\binom{k+j}{k}P'(n''=k)}{\binom{N}{k}}\right]\binom{N}{k+j} \tag{3.5}$$

for $j=1,.....,\overline{k}-\underline{k}$

where, $\underline{k}=\min\left\{k:\binom{k}{2}\geq n'\right\}$; and $\overline{k}=\min\{n'2,N\}$ (3.6)

We propose an estimator based on these $\binom{n''}{2}$ distinct 2-tuples and call it as full-information estimator, which has been defined as

$$\tilde{\bar{T}}_{s\left(\substack{n''\\2}\right)}=\frac{1}{\binom{n''}{2}}\sum_{k=1}^{\binom{n''}{2}}t_k \tag{3.7}$$

The above proposed estimator could be viewed as a Rao Blackwellization of the estimator $\hat{\bar{T}}^*_{s_{n'}}$ based on srswor of 2-tuples from U' given the sufficient statistics $s\binom{n''}{2}$, that is, $\tilde{\bar{T}}_{s\left(\substack{n''\\2}\right)} = E^{S_n|s\left(\substack{n''\\2}\right)}\left(\hat{\bar{T}}^*_{S_{n'}}\right)=\frac{1}{\binom{n''}{2}}\sum_{k=1}^{\binom{n''}{2}}t_k$. Moreover,

$V\left[\tilde{\bar{T}}_{s\left(\substack{n''\\2}\right)}\right]\leq V\left[\hat{\bar{T}}^*_{S_{n'}}\right]$.

We shall now state the expression for variance of the proposed estimator $\tilde{\bar{T}}_{s\left(\substack{n''\\2}\right)}$ and its relative efficiency in the following theorem.

Let us consider the three conditioning stages as follows:

Stage I: Decide about the number n'' of distinct units from *U*.

Stage II: Select a srswor of n'' units from *U*. Given these n'' units a total of $\binom{n''}{2}$ 2-tuples are generated.

Stage III: A srswor of size n' 2-tuples are selected from the $\binom{n''}{2}$ 2-tuples obtained at stage II.

Under these conditioning stages, we shall prove the following theorem :

Theorem 3.1. Under the above 3 stages of sample selection of $s_{n'}$, the following results hold :

1. The estimator $\bar{\bar{T}}_{s_{n'}} \mid n'', \mathrm{II} = \frac{1}{n'}\sum_{s_{n'}} t_k \mid n''$, *II and the proposed full-information estimator* $\tilde{\bar{T}}_{s\binom{n''}{2}}$ *estimate* $\bar{T}_{U'}$ *unbiasedly for all values of* n''.

2. $V\left(\bar{\bar{T}}_{s_{n'}} \mid n'', \mathrm{II}\right) \geq V_2\left(\bar{\bar{T}}_{s\binom{n''}{2}} \mid n''\right) =$

$$A_1 \sum_{U'} t_k^2 + A_2\left\{\sum\nolimits_{U'}^{(1)} t_k t_{k'} + \sum\nolimits_{U'}^{(2)} t_k t_{k'} + \sum\nolimits_{U'}^{(3)} t_k t_{k'}\right\} + A_3 \sum\nolimits_{U'}^{(4)} t_k t_{k'}$$

Where $A_1 = \frac{1}{\binom{n''}{2}\binom{N}{2}}\left(1 - \frac{n''(n''-1)}{N(N-1)}\right)$; $A_2 = \frac{2}{\binom{n''}{2}\binom{N}{2}}\left(\frac{(n''-2)}{(N-2)} - \frac{n''(n''-1)}{N(N-1)}\right)$;

and $A_3 = \frac{2}{\binom{n''}{2}\binom{N}{2}}\left(\frac{(n''-2)(n''-3)}{(N-2)(N-3)} - \frac{n''(n''-1)}{N(N-1)}\right)$

3. $V\left(\bar{\bar{T}}_{s_{n'}}\right) = E_1(A_1 + B_1)\sum_{U'} t_k^2 + E_1(A_2 + B_2)\left\{\sum\nolimits_{U'}^{(1)} t_k t_{k'} + \sum\nolimits_{U'}^{(2)} t_k t_{k'} + \sum\nolimits_{U'}^{(3)} t_k t_{k'}\right\}$

$+\ E_1(A_3 + B_3)\sum\nolimits_{U'}^{(4)} t_k t_{k'} = C$ (say) *Where A's are as in (2), and B's are defined as* $B_1 = \frac{1}{\binom{N}{2}}\left[\frac{1}{n'} - \frac{1}{\binom{n''}{2}}\right]$; $B_2 = 2\left[\frac{1}{n'} - \frac{1}{\binom{n''}{2}}\right] . \frac{1}{\binom{n''}{2} - 1} \frac{\binom{n''}{3}}{\binom{n''}{2}\binom{N}{3}}$;

$$B_3 = 2\left[\frac{1}{n'} - \frac{1}{\binom{n''}{2}}\right] . \frac{1}{\binom{n''}{2} - 1} \frac{\binom{n''}{4}}{\binom{n''}{2}\binom{N}{4}}$$

And

4. $V\left[\tilde{\bar{T}}_{s\binom{n''}{2}}\right] = E_1 V_2\left(\tilde{\bar{T}}_{s\binom{n''}{2}}\right) = E_1(A_1)\sum_{U'} t_k^2 + E_1(A_2)\left\{\sum\nolimits_{U'}^{(1)} t_k t_{k'} + \sum\nolimits_{U'}^{(2)} t_k t_{k'} + \sum\nolimits_{U'}^{(3)} t_k t_{k'}\right\}$

$+\ E_1(A_3)\sum\nolimits_{U'}^{(4)} t_k t_{k'} = D$ (say).

5. Relative efficiency (RE) of using $\tilde{\bar{T}}_{s\binom{n''}{r}}$ *over* $\bar{\bar{T}}_{s_{n'}}$ *shall be given by*

$$\mathrm{RE} = \mathrm{C/D} \tag{3.8}$$

where, C and D are given in (3) and (4) and E_1 *corresponds to the probability distribution of distinct units* n'' *in the random set* $A_{n''}$ *obtained in corollary 2.1.* (proof of the theorem has been given in Appendix).

The calculations of RE could easily be performed by using the probability distribution of n'' obtained in corollary 2.1.

In case of simple random sampling without replacement design D', the full-information estimator $\tilde{\bar{T}}_{s\binom{n''}{2}}$ becomes $s_y^2 = \frac{1}{n''-1}\sum_1^{n''}(y_i - \bar{y})^2$ as an estimator of s_y^2 when $t_k(y) = \frac{1}{2}(y_i - y_j)^2$, $k = (i,j) = 1,2\ldots\binom{N}{2}$;

$c_{yz} = \frac{1}{n''-1}\sum_1^{n''}(y_i - \bar{y})(z_i - \bar{z})$ of C_{yz}

when $v_{YG} = \sum_{i<j=1}^{n''}\frac{1}{\pi_{ij}}(\pi_i\pi_j - \pi_{ij})\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$,

where $\pi_i = \pi_j = \frac{n''}{N} \cdot \pi_{ij} = \frac{n''(n''-1)}{N(N-1)}$ (since the present design D' is srswor) an estimator of V_{YG}.

A re-accounting and a note on the application of the above results is required at this stage. In the theorem 3.1 it has been shown that the full-information estimator $\tilde{\bar{T}}_{s\binom{n''}{2}}$ is superior to the one based on srswor sample $s_{n'}, \bar{\bar{T}}_{s_{n'}}$ Therefore, this suggests to draw n'' distinct units from U to generate $\binom{n''}{2}$ 2-tuples to calculate $\tilde{\bar{T}}_{s\binom{n''}{2}}$. One simplest consideration for deciding n'' could be the cost function $C = C_0 + c.n''$, where C_0 being the overhead cost, and c the cost of observing a unit in U. This cost function is reasonable when the major cost is born in observing the units and the cost of calculating the estimate is comparatively negligible. This consideration suggests us to select as large n'' distinct units from U as permitted by $n'' \le \frac{C - C_0}{c}$.

To sum up the above development, it is suggested that we shall initially select a sample of n'' units from U. Next, we consider all 2-tuples generated by it, $s\binom{n''}{2}$, and estimate $\bar{T}_{U'}$ by utilizing t-measures on generated sample, $s\binom{n''}{2}$, of 2-tuples. For the suggested full information estimator we need to know the partition of S', ppswr sample space, on which each sample reduces into a given $A_{n''}$ and also the conditional probabilities on this partition. Theorem 2.1 gives the required probability distribution of $A_{n''}$ in the present setting. This theorem also helps in identifying s_n's resulting into a fixed $A_{n''}$ starting with $n'' = 2$ and progressing recursively, instead of checking every sample s_n among N'^n samples in S' one by one that gives the same $A_{n''}$.

4. Empirical study

The proposed estimators were applied to estimate the population variance of imports on a real population C0124 of 124 countries consisting of data on 1983 import figures (in millions of U.S. dollars) and 1982 figures on gross national product (in tens of millions of U.S. dollars) (Sarndal, Swensson, Wretman (1991). The correlation between y(IMP) and x(GNP) and $t(\mathbf{y})$ and $t(\mathbf{x})$ are both high; 0.91 and 0.94 respectively. On the basis of computer simulations we have obtained the following results, for the efficiencies of discussed estimators as compared to the estimator $\widehat{T}_{s_n}$.

Table 1. Relative Efficiencies

Estimator of S_y^2	$\widehat{\bar{T}}_{s_n}$	$\widehat{\bar{T}}^{*}_{s'_n}$	$\tilde{\bar{T}}_{s\binom{n''}{2}}$	$s_{y\ conv.}^2$
Design: (ppswr)	1	2.11 2.09	4.59	1.96
Design: (srswr)	1		3.72	1.66

$RE = V\left(\widehat{T}_{s_n}\right)/V(.)$

In these calculations, the conventional estimator of s_y^2 has been taken as $s_{yconv}^2 = \frac{1}{n}\sum_{s_n}\frac{(y_i-\bar{y})^2}{(N-1)P_i}$ which is unbiased for S_y^2. Its variance has been given by $V\left(s_{yconv}^2\right) = \left\{\sum_U \frac{(y_i-\bar{y})^2}{(N-1)P_i} - S_y^4\right\}$. The results in table 1 show clearly that the proposed full information estimator claims overall superiority. Other estimators of variance available in literature have not been included in this study since these and proposed estimator are based on different designs, therefore, non comparable.

Appendix

PROOF OF THEOREM 2.1

Let the set B of M fixed units from U be given by $B = \left\{U_{l_1}, U_{l_2}, \ldots., U_{l_M}\right\}$, U_{l_j} 's being taken from U, so that $2 \le M \le N$. The ppswr sample, s_n, gives the random set $A_{n''} = \left\{U_{i_1}, U_{i_2}, \ldots., U_{i_n}\right\}$, $|A_{n''}| = n'' \le M$. Under the design D'

$$P'[A_{n''} \subset B] = \sum_{S'_B} P'\left(s_n^{(B)}\right) \tag{A.1}$$

The set $A_{n''}$ is a random set which is one set among the class of all subsets of *U*. The size of $A_{n''}$, *i.e.* $|A_{n''}|$ would range from $\underline{k} = 2$ to $\overline{k} = \min\{2n, N\}$. Further,

$$P'[A_{n''} \subset B] = \sum_{\{A_{n''}: A_{n''} \subset B\}} P'(A_{n''}) = \sum_{k=\underline{k}}^{\overline{k}} \sum_{\substack{\{A_{n''}: |A_n| = k, \\ A_{n''} \subset B\}}} P'(A_{n''}) \tag{A.2}$$

On comparing (14) and (15), we get,

$$\sum_{k=\underline{k}}^{\overline{k}} \sum_{\substack{\{A_{n''}: |A_n| = k, \\ A_{n''} \subset B\}}} P'(A_{n''}) = \sum_{S'_B} P'\left(s_n^{(B)}\right) \tag{A.3}$$

for every set *B,* so that $|B| = M \geq \underline{k}$.

If we solve (16) for $P'(A_{n''})$ starting with every *B* of size $\underline{k}$*, i.e.,* $|B| = \underline{k}(M = 2)$, one gets probabilities of every sets $A_{n''}$ of size 2.

Next, for some fixed *B* of size $\underline{k} + 1(2+1)$, we have

$$P'(A_{n''})|_{\substack{A_{n''}=B, \\ |B|=2+1}} = \sum_{S'_B} P'\left(s_n^{(B)}\right) - \sum_{\substack{\{A_{n''}: |A_n| = r, \\ A_{n''} \subset B\}}} P'(A_{n''}) \tag{A.4}$$

where the second term on right hand expression comes from probabilities $P'(A_{n''})$ at the previous step. The probabilities of all such $A_{n''}$'s each of size (2 + 1) could be calculated simply by taking variation over all subsets *B* of size 2 + 1 . Proceeding similarly for $j = 2$ to $\overline{k} = \min\{2n, N\}$, we get a complete probability distribution of $A_{n''}$. More generally, for $\underline{k} + j$ and for fixed *B,* $|B| = \underline{k} + j$

$$P'(A_{n''})|_{\substack{A_{n''}=B, \\ |B|=\underline{k}+j}} = \sum_{S'_B} P'\left(s_n^{(B)}\right) - \sum_{\substack{\{A_{n''}: |A_n| = \underline{k}+j-1, \\ A_{n''} \subset B\}}} P'(A_{n''}) \tag{A.5}$$

$$j = 1, \ldots, \overline{k} - \underline{k}$$

Varying *B,* each of size $(\underline{k} + j)$, one gets the probabilities $P'(A_{n''})$ so that $A_{n''} = B$. Note that the second term on the right side of the above expression comes from probabilities $P'(A_{n''})$ obtained at the previous step. The above recursive method enables one to get the complete probability distribution of $A_{n''}$. This completes the proof.

The advantage of the recursive procedure explained in Theorem 2.1 is that one needs only to calculate the probabilities of all such ppswr samples of size *n*, $s_n^{(M)}$'s drawn from U'_M, the space of all 2-tuples made of units from fixed set *B*, and proceed with $M = \underline{k}$. For variations in *B* of size $\underline{k}$, one gets $P'(A_{n''})$ with $B = A_{n''}$. These probabilities are then used to calculate $P'(A_n'')$, $|A_{n''}| = \underline{k}+1$ for variations in *B*, $B = A_{n''}$ so that $|B| = \underline{k}+1$. Proceeding similarly, one gets the complete probability distribution of $A_{n''}$.

ILLUSTRATION OF THEOREM 2.1: Consider a ppswr design, $D(U,S,P)$, $U = \{U_1, U_2, U_3, U_4\}$; and *S* consists of $4^3 = 64$ ppswr samples of size $n = 3$; the sizes of the units U_1, U_2, U_3, U_4 are 1.00, 2.00, 3.00, 4.00 respectively, giving their selection probabilities 0.1, 0.2, 0.3 and 0.4 respectively. Let us now construct a design $D'(U', S', P')$ so that

$$U' = \{\mathbf{U}'_1 = (U_1, U_2), \mathbf{U}'_2 = (U_1, U_3), \mathbf{U}'_3 = (U_1, U_4), \mathbf{U}'_4 = (U_2, U_3),$$

$$\mathbf{U}'_5 = (U_2, U_4), \mathbf{U}'_6 = (U_3, U_4)\};\ S'$$

is sample space of ppswr samples of size $n = 3$ having $(N')^n = 6^3 = 216$ samples, where, $N' = \binom{N}{2} = 6$, $n = 3$; and the induced probabilities, are $P'(\mathrm{U}'_1) = 0.0472 = P'_1$, $P'(\mathrm{U}'_2) = 0.0762 = P'_2$, $P'(\mathrm{U}'_3) = 0.111 = P'_3$, $P'(\mathrm{U}'_4) = 0.161 = P'_4$, $P'(\mathrm{U}'_5) = 0.233 = P'_5$, $P'(\mathrm{U}'_6) = 0.3714 = P'_6$. Note $\underline{k} = 2$ and $\overline{k} = \min\{nr, N\} = \min\{6, 4\} = 4$. Now starting with $M = \underline{k} = 2$ for the set $B = \{1, 2\}$, the theorem gives,

$$P(A_{n''} = \{1,2\}) = P'\left(s_n^{(B)}\right) = P'\{(1,2),(1,2),(1,2)\}$$

$$= \{P'(\mathbf{U}'_1)\}^3 = P_1'^3 = 1.052 \times 10^{-4}$$

similarly, we get $P(A_{n''} = \{1,3\}) = P_2'^3 = 4.425 \times 10^{-4}$;

$P(A_{n''} = \{1,4\}) = P_3'^3 = 1.3676 \times 10^{-3}$; $P(A_{n''} = \{2,3\}) = P_4'^3 = 4.173 \times 10^{-3}$; $P(A_{n''} = \{2,4\}) = P_5'^3 = 0.0126$; $P(A_{n''} = \{3,4\}) = P_6'^3 = 0.0512.$. Next, consider sets *B* of size $\underline{k}+1 = 3$, we get the probabilities of sets $A_{n''}$ of size 3. Let us fix *B* = {1, 2, 3}, we have U'_B = {(1, 2), (1, 3), (2, 3)}; sample space under ppswr from U'_B would be given by

$S'_B = \{\mathrm{U}'_1, \mathrm{U}'_1, \mathrm{U}'_1\}, \{\mathrm{U}'_2, \mathrm{U}'_2, \mathrm{U}'_2\}, \{\mathrm{U}'_4, \mathrm{U}'_4, \mathrm{U}'_4\}, \{\mathrm{U}'_1, \mathrm{U}'_1, \mathrm{U}'_2\}, \{\mathrm{U}'_1, \mathrm{U}'_1, \mathrm{U}'_4\}$ $\{\mathrm{U}'_2, \mathrm{U}'_2, \mathrm{U}'_1\}, \{\mathrm{U}'_2, \mathrm{U}'_2, \mathrm{U}'_4\}, \{\mathrm{U}'_4, \mathrm{U}'_4, \mathrm{U}'_1\}, \{\mathrm{U}'_4, \mathrm{U}'_4, \mathrm{U}'_2\}$ all with three permutations, and $\{U_1', U_2', U_4'\}$ with 3 permutations. $\sum_{S_B'} P'\left(s_n^{(B)}\right)$ becomes $P_1'^3 + P_2'^3 + P_4'^3 + 3\left\{P_1'^2 (P_2' + P_4') + P_2'^2 (P_1' + P_4') + P_4'^2 (P_1' + P_2')\right\} + 3!(P_1'.P_2'.P_4')$.

The result of the theorem gives,

$$P'(\{1,2,3\}) = \sum_{S_B'} P'\left(s_n^{(B)}\right) - \sum_{\substack{\{A_{n''}:|A_{n''}|=2,\\ A_{n''}\subset B\}}} P'(A_{n''}) \tag{A.6}$$

Note $P'(A_{n''}), |A_{n''}| = 2$, have been calculated at the previous step. This gives,

$$P'(\{1,2,3\}) = 3\left\{P_1'^2 (P_2' + P_4') + P_2'^2 (P_1' + P_4') + P_4'^2 (P_1' + P_2')\right\} + 6(P_1'.P_2'.P_4')$$

$$= 0.018282.$$

similarly $P'(\{1,2,4\}) =$ 0.04576, $P'(\{1,3,4\}) =$ 0.12126, $P'(\{2,3,4\}) = 0.38035$. Having obtained the probabilities of sets $A_{n''}$ of size 3, we would now obtain $A_{n''}$ = {1, 2, 3, 4}, fix B = {1, 2, 3, 4}. Note that $S_B' = S'$ since $A_{n''} = U$, therefore, $\sum_{S_B'} P'\left(s_n^{(B)}\right) = 1$. The result of the theorem gives,

$$P'(\{1,2,3,4\}) = \sum_{S_B'} P'\left(s_n^{(B)}\right) - \sum_{\substack{\{A_{n''}:|A_{n''}|=3,\\ A_{n''}\subset B\}}} P'(A_{n''}) - \sum_{\substack{\{A_{n''}:|A_{n''}|=2,\\ A_{n''}\subset B\}}} P'(A_{n''}) = 0.3645.$$

PROOF OF COROLLARY 3.1. Choose and fix *M* units from *U* and denote them by a subset *B*, $|B| = M$, $1 \le M \le N$ and draw a srswor sample $s_{n'}$ of size n' from U', $P'(s_{n'}) = \dfrac{1}{\binom{\binom{N}{2}}{n'}}$; the corresponding sample space is denoted by S' and the set of distinct units in $s_{n'}$ be denoted by $A_{n''} = \left\{\mathbf{U}_{\mathbf{i}_1}' \cup \mathbf{U}_{\mathbf{i}_2}' \cup ..... \cup \mathbf{U}_{\mathbf{i}_{n'}}'\right\}$ with size $|A_{n''}| = n''$, where $\mathbf{U}_{\mathbf{i}_1}' \cup \mathbf{U}_{\mathbf{i}_2}' \cup ..... \cup \mathbf{U}_{\mathbf{i}_{n'}}'$ gives distinct units which have appeared in $\mathbf{U}_{\mathbf{i}}'$'s in $s_{n'}$. In U' there are $\binom{M}{2}$, 2-tuples consisting entirely of the units in *B,* we denote them by $U_M', U_M' \subset U'$. Then

$$P'[A_{n''} \subset B] = \frac{\binom{M'}{n'}}{\binom{N'}{n'}} \tag{A.7}$$

since n' r-tuples in $s_{n'}$ could be chosen in $\binom{M'}{n'}$ ways from U'_M out of total ways $\binom{N'}{n'}$, where $M' = \binom{M}{2}$, $N' = \binom{N}{2}$, $|B| = M$ and $|A_{n''}| \leq M$. Let the probabilities of $A_{n''}$ having k distinct units be denoted by $P'(n'' = k)$ for $k = \underline{k}, \underline{k}+1, \ldots., \overline{k}$. We shall observe that $P'(A_{n''})$ for $A_{n''} \subset U$, depends only on the size of $A_{n''}$, *i.e.,* $|A_{n''}|$, not on which labels constitute $A_{n''}$. In other words, the distribution of $A_{n''}$ is invariant under the permutations of U_i's. Thus for some $A_{n''} \subset U$ such that $|A_{n''}| = k$, we have

$$P'(A_{n''}) = \frac{P'(n'' = k)}{\binom{N}{k}} \tag{A.8}$$

But, $P'[A_{n''} \subset B] = \sum_{\{T : T \subset B\}} P'(T), \quad |B| = M$

$$= \sum_{k=\underline{k}}^{\overline{k}} \binom{M}{k} P'(A_{n''})|_{|A_{n''}|=k}$$

From (21) we get $P'[A_{n''} \subset B] = \sum_{k=\underline{k}}^{\overline{k}} \binom{M}{k} \frac{P'(n'' = k)}{\binom{N}{k}}$ (A.9)

On comparing (20) and (22) we get

$$\sum_{k=\underline{k}}^{\overline{k}} \frac{\binom{M}{k}}{\binom{N}{k}} P'(n'' = k) = \frac{\binom{M'}{n'}}{\binom{N'}{n'}} \qquad \forall M \geq \underline{k} \tag{A.10}$$

These equations can be recursively solved for $P'(n'' = k)$ starting with $M = \underline{k}$, we get

$$P'(n'' = \underline{k}) = \frac{\binom{\binom{\underline{k}}{2}}{n'} \binom{N}{\underline{k}}}{\binom{N'}{n'}}$$

$$P'(n'' = \underline{k}+1) = \left[\frac{\binom{\binom{\underline{k}+1}{2}}{n'}}{\binom{\binom{N}{2}}{n'}} - \frac{(\underline{k}+1) P'(n'' = \underline{k})}{\binom{N}{\underline{k}}} \right] \binom{N}{\underline{k}+1}$$

$$P'(n''=\underline{k}+j)=\left[\frac{\left(\begin{pmatrix}\underline{k}+j\\ r\\ n'\end{pmatrix}\right)}{\begin{pmatrix}N\\ r\\ n'\end{pmatrix}}-\sum_{k=\underline{k}}^{\underline{k}+j-1}\frac{\binom{k+j}{k}P'(n''=k)}{\binom{N}{k}}\right]\binom{N}{\underline{k}+j}$$

For $j=1,....,\bar{k}-\underline{k}$.

This is the desired distribution of number of distinct units n'' in a srswor $s_{n'}$ taken from U' under the design $D'(U',S',P')$.

ILLUSTRATION OF COROLLARY 2.1

Let N = 5, n' = 3. Then $\underline{k}$ = 3 and $\bar{k}$ = 5 and the equation (7) and (8) reduces to

$$P'(n''=3)=\binom{5}{3}\left(\begin{pmatrix}3\\2\\3\end{pmatrix}\right)\Big/\left(\begin{pmatrix}5\\2\\3\end{pmatrix}\right)=\frac{1}{12};\qquad \frac{1}{\binom{5}{4}}P'(4)+\frac{4}{\binom{5}{3}}P'(3)=\left(\begin{pmatrix}4\\2\\3\end{pmatrix}\right)\Big/\left(\begin{pmatrix}5\\2\\3\end{pmatrix}\right)=\frac{1}{6}$$

and $P'(5)+P'(4)+P'(3)=1$.

Thus $P'(3)=\frac{1}{12}$; $P'(4)=\left(\frac{1}{6}-\frac{4}{10}.\frac{1}{12}\right).5=\frac{2}{3}$ and $P'(5)=\frac{1}{4}$

PROOF OF THEOREM 3.1

1. The conditional expectation of $\bar{\bar{T}}_{s_{n'}}$, given n'' and II is

$$E_3\left(\bar{\bar{T}}\mid n'',\text{II}\right)=\frac{1}{\binom{n''}{2}}\sum_{k=1}^{\binom{n''}{2}}t_k=\tilde{\bar{T}}_{s\binom{n''}{2}}$$

and $E_2E_3\left(\bar{\bar{T}}_{s_{n'}}\mid n'',\text{II}\right)=E_2\left(\tilde{\bar{T}}_{s\binom{n''}{2}}\right)$

$$=\frac{1}{\binom{n''}{2}}\sum_{k=1}^{\binom{n''}{2}}E_2\left(t_k\right)=\sum_{U'}t_kP\left(\mathbf{U'_k}=\left(U_i,U_j\right)\right)$$

$$=\sum_{U'}t_k\cdot\frac{2}{N\left(N-1\right)}=\bar{T}_{U'}$$

This shows that $\bar{\bar{T}}_{s_{n'}}$, and $\tilde{\bar{T}}_{s\binom{n''}{2}}$ are unbiased estimators of $\bar{T}_{U'}$ for all values of n''.

2. Now $V\left(\overline{\overline{T}}_{s_{n'}} \mid n'', \mathrm{II}\right) = V_2 E_3(\cdot \mid n'', II) + E_2 V_3(\cdot \mid n'', II) \geq V_2 E_3(\cdot \mid n'', II)$

$= V_2\left(\tilde{\tilde{T}}_{s\binom{n''}{2}}\right)$

Showing $\tilde{\tilde{T}}_{s\binom{n''}{2}}$ is superior to $\overline{\overline{T}}_{s_{n'}} \mid n'', \mathrm{II}$. Let us now consider

$$V_2\left(\tilde{\tilde{T}}_{s\binom{n''}{2}}\right) = \frac{1}{\left\{\binom{n''}{2}\right\}^2}\left[E_2\left(\sum_{k=1}^{\binom{n''}{2}} t_k\right)^2 - E_2^2\left(\sum_{k=1}^{\binom{n''}{2}} t_k\right)\right] \tag{A.11}$$

Here, $E_2\left(\sum_1^{\binom{n''}{2}} t_k\right)^2$ calculates to be

$$\frac{n''(n''-1)}{N(N-1)}\sum_{U'} t_k^2 + 2\frac{n''(n''-1)(n''-2)}{N(N-1)(N-2)}\left\{\sum_{U'}^{(1)} t_k t_{k'} + \sum_{U'}^{(2)} t_k t_{k'} + \sum_{U'}^{(3)} t_k t_{k'}\right\}$$

$$+2\frac{n''(n''-1)(n''-2)(n''-3)}{N(N-1)(N-2)(N-3)}\sum_{U'}^{(4)} t_k t_{k'}$$

where

$\sum_{U'}^{(1)} = \sum_{\substack{k \neq k' \\ k=(i,j),\, k'=(i,l) \\ i<j<l}}$; $\sum_{U'}^{(2)} = \sum_{\substack{k \neq k' \\ k=(i,j),\, k'=(j,l) \\ i<j<l}}$; $\sum_{U'}^{(3)} = \sum_{\substack{k \neq k' \\ k=(i,j),\, k'=(k,j) \\ i<k<j}}$;

$\sum_{U'}^{(4)} = \sum_{\substack{k \neq k' \\ k=(i,j),\, k'=(k,l) \\ i<j<k<l \\ i \neq j \neq k \neq l}}$; and $\left\{E_2\left(\sum_1^{\binom{n''}{2}} t_k\right)\right\}^2$ calculates to be $\left(\frac{n''(n''-1)}{N(N-1)}\right)^2$

$$\left\{\sum_{U'} t_k^2 + 2\sum_{U'}^{(1)} t_k t_{k'} + 2\sum_{U'}^{(2)} t_k t_{k'} + 2\sum_{U'}^{(3)} t_k t_{k'} + 2\sum_{U'}^{(4)} t_k t_{k'}\right\}.$$

On putting these values in (24), we get

$$V_2\left(\tilde{\tilde{T}}_{s\binom{n''}{2}}\right) = A_1\sum_{U'} t_k^2 + A_2\left\{\sum_{U'}^{(1)} t_k t_{k'} + \sum_{U'}^{(2)} t_k t_{k'} + \sum_{U'}^{(3)} t_k t_{k'}\right\} + A_3\sum_{U'}^{(4)} t_k t_{k'}$$

Where $A_1 = \frac{1}{\binom{n''}{2}\binom{N}{2}}\left(1 - \frac{n''(n''-1)}{N(N-1)}\right)$; $A_2 = \frac{2}{\binom{n''}{2}\binom{N}{2}}\left(\frac{(n''-2)}{(N-2)} - \frac{n''(n''-1)}{N(N-1)}\right)$;

$A_3 = \frac{2}{\binom{n''}{2}\binom{N}{2}}\left(\frac{(n''-2)(n''-3)}{(N-2)(N-3)} - \frac{n''(n''-1)}{N(N-1)}\right)$;

3. Next, consider

$$V_3\left[\overline{\overline{T}}_{s_{n'}} \mid n''\right]=\left[\frac{1}{n'}-\frac{1}{\binom{n''}{2}}\right]\frac{1}{\binom{n''}{2}-1}\left[\sum_1^{\binom{n''}{2}} t_k^2-\frac{1}{\binom{n''}{2}}\left(\sum_1^{\binom{n''}{2}} t_k\right)^2\right]$$

Taking E_2 on $V_3\left[\overline{\overline{T}}_{s_{n'}}\right]$, some algebraic simplification gives

$$E_2 V_3\left[\overline{\overline{T}}_{s_{n'}}\right]=B_1\sum_{U'} t_k^2-B_2\left\{\sum_{U'}^{(1)} t_k t_{k'}+\sum_{U'}^{(2)} t_k t_{k'}+\sum_{U'}^{(3)} t_k t_{k'}\right\}-B_3\sum_{U'}^{(4)} t_k t_{k'} \quad \text{(A.12)}$$

Where $B_1=\frac{1}{\binom{N}{2}}\left[\frac{1}{n'}-\frac{1}{\binom{n''}{2}}\right]$; $B_2=\left[\frac{1}{n'}-\frac{1}{\binom{n''}{2}}\right].\frac{1}{\binom{n''}{2}-1}.\frac{\binom{n''}{3}}{\binom{n''}{2}\binom{N}{3}}$;

$$B_3=2.\left[\frac{1}{n'}-\frac{1}{\binom{n''}{2}}\right].\frac{1}{\binom{n''}{2}-1}.\frac{\binom{n''}{4}}{\binom{n''}{2}\binom{N}{4}}$$

So far we have been treating n'' as a quantity fixed in advance, but in actual practice we fix n' and select n' 2-tuples from $\binom{N}{2}$ 2-tuples by srswor. For any given sample $s_{n'}$, we get n'' distinct units of *U*. Thus n'' is actually a random variable.

The distribution of n'' has already been calculated in corollary 2.1. by recursive method. We now have,

$$V\left(\overline{\overline{T}}_{s_{n'}}\right)=E_1E_2V_3(\cdot)+E_1V_2E_3(\cdot)+V_1E_2E_3(\cdot)$$

$$=E_1(A_1+B_1)\sum_{U'} t_k^2+E_1(A_2+B_2)\left\{\sum_{U'}^{(1)} t_k t_{k'}+\sum_{U'}^{(2)} t_k t_{k'}+\sum_{U'}^{(3)} t_k t_{k'}\right\}$$

$$+E_1(A_3+B_3)\sum_{U'}^{(4)} t_k t_{k'}=C \text{ (say)} \quad \text{(A.13)}$$

4. From (25), we get

$$V\left(\tilde{\tilde{T}}_{s\binom{n''}{2}}\right)=E_1V_2\left(\tilde{\tilde{T}}_{s\binom{n''}{2}}\right)$$

$$=E_1(A_1)\sum_{U'} t_k^2+E_1(A_2)\left\{\sum_{U'}^{(1)} t_k t_{k'}+\sum_{U'}^{(2)} t_k t_{k'}+\sum_{U'}^{(3)} t_k t_{k'}\right\}$$

$$+E_1(A_3)\sum_{U'}^{(4)} t_k t_{k'}=D \text{ (say)} \quad \text{(A.14)}$$

5. The relative efficiency (RE) of the estimator $\tilde{\bar{T}}_{s\left(\frac{n''}{2}\right)}$ has been given by

$$RE = \frac{V\left(\bar{\bar{T}}_{s_{n'}}\right)}{V\left(\tilde{\bar{T}}_{s\left(\frac{n''}{2}\right)}\right)} = C/D$$

REFERENCES

[1] CASSEL, C.M., SARNDAL, C.E., and WRETMAN, J.H. (1977). *Foundations of Inference in Survey Sampling.* New York: Wiley.

[2] CHAUDHRI, A.(1978). On Estimating the Variance of a Finite Population. *Metrika* **25**, 65–76.

[3] DATTA, G.S. and GHOSE, M.(1993).Bayesian Estimation of Finite Population Variances with Auxiliary Information. *Sankhya, Ser.* B,**55**, 156–170.

[4] DATTA and TIWARI (1991).Bayesian Estimation of Finite Population Variances with Auxiliary Information. *Sankhya, Ser.* B,**55**, 156–170.

[5] GHOSH, M. and LAHIRI, P.(1987).Robust Empirical Bayes Estimation of Variance From Stratified Samples. *.J. Amer. Statist, Assoc.,***82***,* 1153–1162.

[6] GHOSE, M. and MEEDEN, G.(1983).Estimation of the Variance in Finite Population Sampling. *Sankhya, Ser. B,**45**,* 362–375.

[7] GHOSE, M. and MEEDEN, G.(1984).A new Bayesian analysis of a random effect model. *J.R. Statist. Soc.,B***46***,* 474–482.

[8] ERICSON, W.A. (1969).Subjective Bayesian Models in Sampling Finite Populations(With discussion). *Journal of Royal Statistical Society, Ser.* B**31**, 195–233.

[9] GODAMBE, V.P.(1955). A unified theory of sampling from finite populations. *J. R. Statist. Soc. B,* **17**, 208–278.

[10] HANURAV, T.V.(1966). Some aspect of unified sampling theory. *Sankhya, Ser. A,* **28**, 175–204.

[11] LAHIRI, P. and TIWARI, R.C.(1991). NonParametric Bayes and Empirical Bayes Estimation of Variances from Stratified Sampling. *Sankhya, Ser.* S,**52**, 105–118.

[12] LIU, T.P. (1974a). Bayes Estimation for the Variance of a finite population. *Metrika* **36**, part I, 23–32.

[13] LIU, T.P. (1974b). A General Unbiased Estimator for the Variance of a Finite Population. *Sankhya, Ser. C,* **36**, part I, 23–32.

[14] LIU, T. P. and THOMPSON, M.E. (1983). Properties of Estimators of Quadratic Finite Population Function the Batch Approach. *Annals of Statistics* **11***,* 275–285.

[15] MUKHOPADHYAY, P.(1978). Estimating the Variance of a Finite Population Under a Superpopulation Model. *Metrika* **25***,* 115–122.

[16] MUKHOPADHYAY, P. and BHATTACHARYA, S. (1989). On Estimating the Variance of a Finite Population Under a Superpopulation Model. *Journal of the Indian Statistical Assoc.**27**,* 37–46.

[17] RAJ, DES (1968). *Sampling theory.* Tata McGraw-Hill Publishing Company Ltd. New Delhi.

[18] SWAIN, A.K.P.C. and MISHRA, G.(1994). Estimation of Finite Population Variance Under Unequal Probability Sampling. *Sankhya, Ser. B,***56***,* 374–388.

[19] SARNDAL.C.E., SWENSSON, B. and WRETMAN, J. (1992).*Model Assisted Survey Sampling,* Springer-Verlag,Newyork.

[20] VARDEMAN, S. and MEEDEN, G. (1983). Admissible Estimators in Finite Population Sampling Employing Various Types of Prior Information. *Journal of Statistical Planning Inference, 7,* 329–341.

[21] WU,C. and SITTER, R.R. (2002). Efficient Estimation of Quadratic Finite Population Functions in the Presence of Auxiliary Information. *J. Amer. Statist. Assoc.,**97**,,* 535–543.

[22] ZACKS, S. and SOLOMAN, H. (l981). Bayes and Equivariant Estimators of the Variance of a finite population:Part I, Simple random Sampling. *Commun. Statist-Theory Meth.,* **A10**, 407–426.