Jour. Ind. Soc. Ag. Statistics
55(2), 2002 : 197-208

Resampling-Based Variance Estimation Under Two-Phase Sampling

V. Ramasubramanian, Randhir Singh and Anil Rai
Indian Agricultural Statistics Research Institute, New Delhi-110 012
(Received: May, 2001)

SUMMARY

Two new Bootstrap methods of variance estimation under two-phase sampling have been developed. Conditional inference for variance estimation under two-phase sampling has also been explored. Comparison of the developed methods vis-a-vis the existing Jackknife method has been discussed under both design-based and conditional design-based settings. The study revealed that the proposed methods compare well with the Jackknife method.

Key words: Bootstrap, Conditional inference, Design-based inference, Jackknife, Simulation.

1. Introduction

Two-phase sampling is commonly adopted for surveys when there is little or no prior information about the population. It is also a viable alternative to simple random sampling when there are expected to be gains from using auxiliary information. Various resampling procedures viz. the Jackknife, the Balanced Repeated Replication (BRR) and the Bootstrap have been suggested in the literature for estimation of variance under two-phase sampling. Schreuder *et al.* [8] performed a simulation study comparing variance estimation techniques for sampling with partial replacement. In the process, they proposed a Bootstrap variance estimator for two-phase sampling though they gave no motivation and no theoretical justification. Biemer and Atkinson [1] suggested a Bootstrap procedure for two-phase sampling wherein the original two-phase sample was replicated a finite number of times to form a 'pseudo-population' and upon this Bootstrap has been applied for estimating variance. A Jackknife procedure has been suggested for the two-phase ratio estimator by Rao and Sitter [5]. Kott [3] discussed Jackknife variance estimation when the first phase sample is used for stratification. Sitter [6] has extensively studied the Schreuder *et al.'s* [8] Bootstrap method and Rao and Sitter's [5] Jackknife method. Rao and Sitter [6] reviewed some of these methods and focussed two-phase sampling for measurement errors in case of stratified design. Fuller [2] suggested a BRR variance estimator for two-phase samples.

The Bootstrap can be employed to estimate variance for a wide range of estimators under two-phase sampling whereas the Jackknife and the BRR procedures are not always easily applicable. As a consequence of this, recently, lot of literature is coming up in this area as cited above. The Bootstrap technique also gained considerable popularity over time due to its simplicity.

It is well known that direct extensions of the standard resampling procedures to problems where the sampling units may not be independent and identically distributed may often lead to misleading inferences. Considering the special structure of the two-phase sampling design, two new Bootstrap methods namely, the two-phase post-stratified Bootstrap method and the two-phase Proportionate Bootstrap method have been developed in this paper which are discussed in the subsequent sections.

2. *Notations and Preliminaries*

Consider a bivariate population (X_i, Y_i), $i = 1, \ldots, N$ from which a sample is drawn by the simple (unstratified) two-phase sampling design wherein simple random sampling without replacement is used at each phase. Let $(n', s^{(1)})$ and $(n, s^{(2)})$ denote the sample size and the set of sampled units at the first and second phase respectively. Let $\bar{x}_{n'}$ and $s_x'^2$ denote the mean and variance of elements of the first phase sample $s^{(1)}$ of size n'. Let $(\bar{x}_n, \bar{y}_n)$ and (s_x^2, s_y^2, s_{xy}) denote the means, variances and covariance of the elements of second phase sample $s^{(2)}$ of size n. Likewise, let $\bar{x}_{n^*}$ and s_x^{*2} denote the mean and variance of the set $(s^{(1)} - s^{(2)})$ of size $(n' - n)$.

Under this set up, a ratio estimator of population mean $\bar{Y}$ can be written as

$$\bar{y}_{dr} = \frac{\bar{y}_n}{\bar{x}_n}\bar{x}_{n'} = \hat{R}\,\bar{x}_{n'}$$

Variance estimation for two-phase samples based on Taylor's linearisation method is described in standard textbooks. The approximate variance estimator of the above ratio estimator is given by, say

$$v_0 = \left(\frac{1}{n} - \frac{1}{n'}\right)\left(s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R}s_{xy}\right) + \frac{1}{n'}s_y^2 \tag{2.1}$$

Rao and Sitter [5] proposed a new linearisation variance estimator that makes more complete use of the sample data, given by, say (ignoring finite population corrections)

$$v_1 = \frac{s_d^2}{n} + 2\hat{R}\frac{s_{dx}}{n'} + \hat{R}^2\frac{s_x'^2}{n'} \tag{2.2}$$

where $s_d^2 = \frac{1}{n-1}\sum_{i \in s^{(2)}} d_i^2$ with $d_i = y_i - \hat{R}x_i, i \in s^{(2)}$ and s_x^2 and s_{dx} are respectively the sample variance of x values and sample covariance of d_i and x_i of the second phase sample $s^{(2)}$.

3. Two-Phase Post-Stratified Bootstrap Method

In the two-phase Post-stratified Bootstrap method, the second phase sample is selected randomly from the first phase sample.

The method is as follows

(i) Draw a simple random resample with replacement of size n' from the first phase sample $s^{(1)}$.

(ii) Post-stratify this resample into two sets, one set having size n_1 (say) consisting of units (with repetitions, if any) that belong to the set $(s^{(1)} - s^{(2)})$, *i.e.* such units that belong to the first phase sample $s^{(1)}$ only but not to the second phase sample $s^{(2)}$ and the other set of size n_2 (say) containing units that belong to the second phase sample $s^{(2)}$. Note that here n_1 and n_2 are random sizes ranging between 0 and n'. Discard the resample for which $n_1 = 0$ or $n_2 = 0$ and select another resample if such is the case, because otherwise it may not be representing a case of two-phase sampling.

(iii) Repeat steps (i) and (ii), say, B times.

The b^{th} Bootstrap resample estimator (b = 1, 2, ..., B) is given by

$$\bar{y}_{dr_1}^b = \frac{\bar{y}_{n_2}^b}{\bar{x}_{n_2}^b}\bar{x}_{n'_r}^b$$

where $$\bar{y}_{n_2}^b = \frac{1}{n_2}\sum_{\substack{i=1 \\ (i \in s^{(2)})}}^{n_2} y_i^b \quad \text{and} \quad \bar{x}_{n_2}^b = \frac{1}{n_2}\sum_{\substack{i=1 \\ (i \in s^{(2)})}}^{n_2} x_i^b$$

$$\bar{x}_{n'_r}^b = \frac{n}{n'}\bar{x}_{n_2}^b + \frac{n'-n}{n'}\bar{x}_{n_1}^b$$

with $$\bar{x}_{n_1}^b = \frac{1}{n_1}\sum_{\substack{i=1 \\ (i \in s^{(1)} - s^{(2)})}}^{n_1} x_i^b$$

Note here that, the sample mean of x from the first phase sample $s^{(1)}$ is computed as a post-stratified estimator, instead of simple mean of all units of $s^{(1)}$. This is because post-stratification is involved in the underlying Bootstrap method and it requires to take the form of the estimator as proposed.

Now, applying the usual Bootstrap method to $\bar{y}_{dr}$ we get the Bootstrap variance estimator as, say

$$v_{B_1} = \frac{1}{B-1}\sum_{b=1}^{B}(\bar{y}^b_{dr_1} - \bar{\bar{y}}_1)^2 \text{ where } \bar{\bar{y}}_1 = \frac{1}{B}\sum_{b=1}^{B}\bar{y}^b_{dr_1} \tag{3.1}$$

For large B, we can replace $\bar{\bar{y}}_1$ by $\bar{y}_{dr}$ itself, so as to get, say, another version of (3.1) given by

$$v_{B_{1a}} = \frac{1}{B-1}\sum_{b=1}^{B}(\bar{y}^b_{dr_1} - \bar{y}_{dr})^2 = E_*(\bar{y}^b_{dr_1} - \bar{y}_{dr})^2 \tag{3.2}$$

where E_* denotes Bootstrap expectation.

For a nonlinear estimator $\theta = g(\bar{Y})$, a Bootstrap variance estimator is obtained readily by replacing $\bar{y}^b_{dr_1}$ and $\bar{y}_{dr}$ by $\hat{\theta}^b_{dr_1} = g(\bar{y}^b_{dr_1})$ and $\hat{\theta}_{dr} = g(\bar{y}_{dr})$ respectively.

Expanding $\bar{y}^b_{dr_1}$ in right hand side (R.H.S.) of (3.2) by Taylor's series, we get

$$\begin{aligned} E_*(\bar{y}^b_{dr_1} - \bar{y}_{dr})^2 &= \hat{R}^2 E_*(\epsilon_1'^2) + \hat{R}^2\left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right)E_*(\epsilon_1^2) + \left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right)^2 E_*(\epsilon_0^2) \\ &\quad - 2\hat{R}^2\left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right)E_*(\epsilon_1\epsilon_1') + 2\hat{R}^2\left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right)E_*(\epsilon_0\epsilon_1') \\ &\quad - 2\hat{R}^2\left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right)^2 E_*(\epsilon_0\epsilon_1) \end{aligned} \tag{3.3}$$

where $\epsilon_0 = \bar{y}^b_{n_2} - \bar{y}_n$; $\epsilon_1 = \bar{x}^b_{n_2} - \bar{x}_n$; $\epsilon_1' = \bar{x}^b_{n_r'} - \bar{x}_{n_r'}$ and $\hat{R} = \frac{\bar{y}_n}{\bar{x}_n}$

Since the Bootstrap sampling design is simple random sampling with replacement, therefore

$$E_*(\epsilon_0^2) = V_*(\bar{y}^b_{n_2}) = V_*^c E_*(\bar{y}^b_{n_2}) + E_*^c V_*(\bar{y}^b_{n_2}) \tag{3.4}$$

where E_*^c, V_*^c denote the conditional Bootstrap expectation and the conditional Bootstrap variance respectively. Now,

$$E_*(\epsilon_0^2) = \left(1 - \frac{1}{n}\right)s_y^2 E_*^c\left(\frac{1}{n_2}\right) \tag{3.5}$$

Following Stephan [10], expand $E_*^c\left(\frac{1}{n_2}\right)$ in an infinite series of inverse factorials to write $E_*^c\left(\frac{1}{n_2}\right) = \sum_{i=1}^{t} u_i + E(R_t(x))$ with $R_t(x) \rightarrow 0$ as $t \rightarrow \infty$

$$u_i = \frac{(i-1)u_{i-1} - k/i}{(n'+i)p},\ i > 1 \text{ and } u_1 = \frac{1-k}{(n'+1)p} \text{ where } k = \frac{n'pq^{n'}}{1-q^{n'}}$$

since $\quad pr(n_2) = \left[\binom{n'}{n_2} p^{n_2} q^{n'-n_2}\right] \Big/ \left(1-q^{n'}\right)$

with $\quad 1 \le n_2 \le n',\ p = \frac{n}{n'},\ q = 1 - \frac{n}{n'}$

For large n', $k \cong 0$ and $E_*^c\left(\frac{1}{n_2}\right) = \frac{n'}{(n'+1)n}$. Hence after simplification we can get (3.5) as

$$E_*\left(\epsilon_0^2\right) = \frac{n'}{(n'+1)}\left(1 - \frac{1}{n}\right)\frac{s_y^2}{n} \text{ and similarly}$$

$$E_*\left(\epsilon_1^2\right) = \frac{n'}{(n'+1)}\left(1 - \frac{1}{n}\right)\frac{s_x^2}{n}$$

$$E_*\left(\epsilon_1'^2\right) = \frac{n-1}{n'(n'+1)} s_x^2 + \frac{n'-n-1}{n'(n'+1)} s_x^{\prime\prime 2}$$

$$E_*(\epsilon_0 \epsilon_1) = \frac{n'}{n'+1}\frac{n-1}{n}\frac{s_{xy}}{n}$$

$$E_*(\epsilon_0 \epsilon_1') = \frac{n-1}{n(n'+1)} s_{xy}$$

$$E_*(\epsilon_1 \epsilon_1') = \frac{n-1}{n(n'+1)} s_x^2$$

Putting these variances and covariances in (3.3) and assuming $(n'-1) \cong n'$ for large $\bar{y}_{dr}$ the linearised version of the Bootstrap estimator of variance can be obtained as

$$\hat{\pi}_h^{(2)} = \hat{R}^2 s_x^2\left[\frac{n}{n'^2} + \left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right)\frac{1}{n} - 2\left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right)\frac{1}{n'}\right] + s_y^2\left[\left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right)^2 \frac{1}{n}\right]$$

$$+ \hat{R}^2 s_x^{\prime\prime 2}\left(\frac{n''}{n'^2}\right) - 2\hat{R} s_{xy}\left[\left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right)^2 \frac{1}{n} - \left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right)\frac{1}{n'}\right] \tag{3.6}$$

Letting $\left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right) \cong 1$, (3.6) reduces approximately to v_0 in (2.1). Hereinafter v_B is referred to as the linearised Bootstrap.

4. Two-Phase Proportionate Bootstrap Method

An important consideration of the Bootstrap method is that the resamples should generally resemble as much as possible the original sample. Hence an alternative method called the two-phase Proportionate Bootstrap method is developed in such a way that the proportion of units belonging to $s^{(2)}$ and $\left(s^{(1)} - s^{(2)}\right)$ appearing in each Bootstrap resample remains the same as in the original sample.

The method is as follows

(i) Draw a simple random sample of size n with replacement from the set $s^{(2)}$ of size n.

(ii) Draw another simple random sample of size $(n' - n) = n''$ (say) with replacement from the set $\left(s^{(1)} - s^{(2)}\right)$ of size $(n' - n)$.

(iii) Repeat the steps (i) and (ii) independently, say, B times.

The b^{th} (b = 1, 2, ..., B) Bootstrap resample estimator is given by

$$\bar{y}^b_{dr_2} = \frac{\bar{y}^b_n}{\bar{x}^b_n} \bar{x}^b_{n'_f}$$

where $$\bar{y}^b_n = \frac{1}{n} \sum_{\substack{i=1 \\ (i \in s^{(2)})}}^{n} y^b_i \; ; \; \bar{x}^b_n = \frac{1}{n} \sum_{\substack{i=1 \\ (i \in s^{(2)})}}^{n} x^b_i$$

$$\bar{x}^b_{n'_f} = \frac{n}{n'} \bar{x}^b_n + \frac{n' - n}{n'} \bar{x}^b_{n'-n}$$

with $$\bar{x}^b_{n'-n} = \frac{1}{n' - n} \sum_{\substack{i=1 \\ (i \in s^{(1)} - s^{(2)})}}^{n'-n} x^b_i$$

Analogous to the method discussed in section 3, the Bootstrap variance estimators for this method are given by, say

$$v_{B_2} = \frac{1}{B-1} \sum_{b=1}^{B} (\bar{y}^b_{dr_2} - \bar{\bar{y}}_2)^2 \quad (4.1)$$

where $$\bar{\bar{y}}_2 = \frac{1}{B} \sum_{b=1}^{B} \bar{y}^b_{dr_2}$$

and
$$v_{B_{2a}} = \frac{1}{B-1}\sum_{b=1}^{B}(\bar{y}_{dr_2}^{b} - \bar{y}_{dr})^2$$
$$= E_*(\bar{y}_{dr_2}^{b} - \bar{y}_{dr})^2$$

Proceeding on similar lines as discussed in section 3 we can see that the linearised version of Bootstrap estimator of variance v_{B_2} in (4.1) from this method comes out to be exactly equal to v_B in (3.6). Note that v_{B_1} in (3.1) has been obtained as v_B after some approximations due to the randomness of the sample sizes involved in the first (previous) method. Thus v_{B_2} also obviously reduces to v_0 in (2.1).

5. Simulation Study

Theoretically we have seen that the two proposed Bootstrap methods yield almost results similar to those from the standard methods of variance estimation. Since these theoretical results are based on asymptotic arguments, their practical applications have been validated by a simulation study in order to examine the properties of the proposed methods vis-a-vis the existing Rao and Sitter's [5] Jackknife and the Sitter's [9] Bootstrap methods.

5.1 Existing Methods

The Rao and Sitter's [5] Jackknife estimator is given by

$$\bar{y}_{dr}(i) = \frac{\bar{y}_{n-1}(i)}{\bar{x}_{n-1}(i)}\,\bar{x}_{n'-1}(i), \ \forall\, i \in s^{(1)},\ i = 1, 2, \ldots, n'$$

where
$$\bar{y}_{n-1}(i) = \begin{cases} \dfrac{n\bar{y}_n - y_i}{n-1} & \text{if } i \in s^{(2)} \\ \bar{y}_n & \text{if } i \in s^{(1)} - s^{(2)} \end{cases}$$

and
$$\bar{x}_{n-1}(i) = \begin{cases} \dfrac{n\bar{x}_n - x_i}{n-1} & \text{if } i \in s^{(2)} \\ \bar{x}_n & \text{if } i \in s^{(1)} - s^{(2)} \end{cases}$$

Also
$$\bar{x}_{n'-1}(i) = \frac{n'\bar{x}_{n'} - x_i}{n'-1} \qquad \forall\, i \in s^{(1)}$$

Rao and Sitter's Jackknife variance estimator is given by

$$v_J = \frac{n'-1}{n'}\sum_{i\in s^{(1)}}(\bar{y}_{dr}(i) - \bar{y}_{dr})^2$$

whose linearised version is given by (hereinafter referred to as the linearised Jackknife)

$$v_{J_1} \cong \left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right)^2 \frac{s_d^2}{n} + 2\left(\frac{\bar{x}_{n'}}{\bar{x}_n}\right) \hat{R} \frac{s_{dx}}{n'} + \hat{R}^2 \frac{s_x'^2}{n'}$$

with notations as defined in section 2 for (2.2).

5.2 Design-Based Inference

To examine the various properties of the proposed methods A = 5000 independent two-phase main samples (with size $n' = 400$ and $n = 80$) have been generated using the following model

$$y_i = \beta x_i + \epsilon_i \sqrt{x_i}$$

where $\epsilon_i \sim N(0, \sigma_e^2)$ independent of x_i which follows generalised gamma distribution with parameters g and h, *i.e.* $x_i \sim G(g, h)$. The model considered for the study has been shown by earlier workers (Royall [7]) to generate the type of population under which the ratio estimator is the best among a wide class of estimators. Note here that β is the regression coefficient of y on x.

The mean, variance and coefficient of variation of x are respectively given by $\mu_x = g\,h$; $\sigma^2_x = g\,h^2$; $C_x = \sigma_x / \mu_x = g^{-1/2}$. Further, the mean and variance of y and the correlation coefficient between (y , x) are respectively given by

$$\mu_y = \beta\mu_x \; ; \; \sigma_y^2 = \beta^2\sigma_x^2 + \mu_x \; \sigma_e^2 \; ; \; \mathrm{corr}(x_i, y_i) = \rho = \beta\frac{\sigma_x}{\sigma_y}$$

Here we fixed $\beta = 1.0$ and $\mu_x = 100$ and choose σ_e and σ_x to match specified values of ρ and C_x. The mean squared error of $\bar{y}_{dr}$ was calculated as

$$\mathrm{MSE}(\bar{y}_{dr}) = \frac{1}{A}\sum_{a=1}^{A} \left(\bar{y}_{dr}^{(a)} - \mu_y\right)^2$$

where $\bar{y}_{dr}^{(a)}$ is the value of $\bar{y}_{dr}$ for the a^{th} simulation run. Also, the simulated mean, mean squared error and the percent relative bias of the specified variance estimator, say v, based on B = 200 Bootstrap resamples from each of the 5000 main samples were calculated as

$$E(v) = \frac{1}{A}\sum_{a=1}^{A} v^a$$

$$\mathrm{MSE}(v) = \frac{1}{A}\sum_{a=1}^{A} \left\{v^{(a)} - \mathrm{MSE}(\bar{y}_{dr})\right\}^2$$

$$\%\mathrm{R.B.}(v) = \frac{E(v) - \mathrm{MSE}(\bar{y}_{dr})}{\mathrm{MSE}(\bar{y}_{dr})} \times 100$$

Table 1. Comparison of the MSEs of the resampling-based variance estimators with that of the usual variance estimator v_0 of two-phase ratio estimator $\bar{y}_{dr}$

Var. estimator	$\rho \downarrow$ \ $C_x \rightarrow$	1.4	0.77	0.33
v_l	0.9	0.23(0.10)	0.70(−1.43)	0.76(2.15)
	0.8	0.82(−1.07)	0.89(−10.53)	0.93(−2.70)
	0.7	0.90(−6.52)	0.94(2.09)	0.96(5.51)
v_{B_1}	0.9	0.27(1.13)	0.96(−0.69)	1.27(1.65)
	0.8	0.95(1.67)	1.11(−10.69)	1.39(−3.05)
	0.7	1.13(−0.72)	1.23(2.97)	1.35(5.45)
$v_{B_{1a}}$	0.9	0.28(2.14)	0.97(0.30)	1.29(2.75)
	0.8	0.97(2.69)	1.09(−9.76)	1.39(−2.06)
	0.7	1.14(−0.46)	1.25(4.08)	1.38(6.55)
v_{B_2}	0.9	0.27(0.59)	1.00(−1.44)	1.23(0.91)
	0.8	0.95(0.86)	1.11(−10.23)	1.35(−2.88)
	0.7	1.09(−2.99)	1.23(1.55)	1.31(4.53)
$v_{B_{2a}}$	0.9	0.27(1.67)	1.01(−0.44)	1.25(1.92)
	0.8	0.96(1.89)	1.10(−9.36)	1.36(−1.81)
	0.7	1.09(−2.07)	1.24(2.56)	1.33(5.62)
v_J	0.9	0.23(3.01)	0.69(0.22)	0.75(2.52)
	0.8	0.82(3.83)	0.87(−8.77)	0.93(−1.54)
	0.7	0.97(0.38)	0.97(4.58)	0.97(6.61)
v_{J_1}	0.9	0.24(0.10)	0.71(−0.87)	0.77(1.89)
	0.8	0.89(0.20)	0.85(−10.18)	0.94(−2.38)
	0.7	1.04(−3.56)	1.02(2.73)	1.01(5.55)

Note:- Values within parentheses indicate percent relative biases

Table 1 presents the values of $MSE(v)/MSE(v_o)$ and per cent relative biases for various variance estimators v for selected values of ρ and C_x. It is clear from the table that the proposed variance estimators v_{B_1}, $v_{B_{1a}}$, v_{B_2} and $v_{B_{2a}}$ are substantially more efficient than v_l for $\rho \geq 0.8$ and become more so as C_x increases. The Bootstrap variance estimators $v_{B_{1a}}$ and $v_{B_{2a}}$ obtained by deviating the resample estimator from the mean of resample means were found better to those deviated from the two-phase ratio estimator i.e. from v_{B_1} and v_{B_2} when their percent relative biases are compared. When the proposed Bootstrap methods are compared with the existing Jackknife method v_J and its linearised version v_{J_l}, the proposed Bootstrap perform almost at par with the Jackknife. When the percent relative biases of v_{B_1} and v_{B_2} are compared v_{B_2} appears to be better than v_{B_1} ; in case of stability both the estimators performed almost equally, as expected, since for large n, the two methods are almost same. However, for smaller n, the two-phase Proportionate Bootstrap method may perform better.

5.3 Conditional Inference

Although inferences through conventional design-based approach may be appropriate at the design stage of the survey but once the sample has been selected and the sample contains "recognisable subsets" then the whole of the sample space may not be the relevant reference set for making inferences (Rao [4]). Hence a conditional design-based approach has been considered further for the study which allows restricting the set of samples used for inference purposes.

To study the conditional properties of the proposed estimators alongwith the existing ones A = 5000 simulated two-phase samples were first ordered based on the values of $(\bar{x}_{n'}/\bar{x}_n)$ i.e. sample ratio of auxiliary character x based on first phase sample $s^{(1)}$ with respect to the second phase sample $s^{(2)}$ and then grouped into 10 successive groups each consisting of 500 samples.

For each group, the simulated conditional mean squared error MSE_c of the two-phase ratio estimator $\bar{y}_{dr}$ was calculated as $MSE_c = \frac{1}{500}\sum_{r=1}^{500}\left(\bar{y}_{dr}^{(r)} - \mu_y\right)^2$.

Also, the conditional mean of a particular variance estimator v was calculated as $E_c(v) = \frac{1}{500}\sum_{r=1}^{500} v^r$. The values of the conditional means of the different variance estimators and the conditional mean squared error were then plotted against the group average of $(\bar{x}_{n'}/\bar{x}_n)$ for the case $C_x = 1.4$; $\rho = 0.8$; n = 80; n′ = 400. For other values of C_x and ρ also we obtained a qualitatively similar plot, hence not presented here.

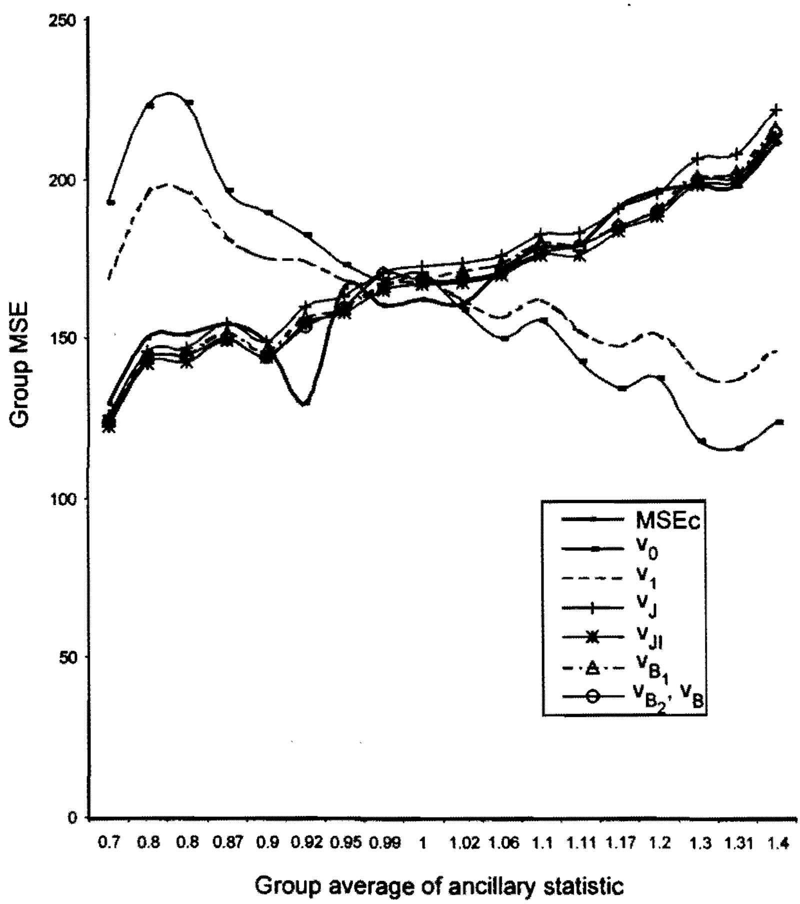

Fig. 1 The performance of the various resampling-based variance estimators under conditional framework

Fig 1 shows that v_{B_1} and v_{B_2} perform well in tracking the conditional mean squared error while v_0 and v_1 lead to significant overestimation of the conditional mean squared error when $(\bar{x}_{n'}/\bar{x}_n) \le 0.8$ and significant underestimation when the samples are negatively unbalanced *i.e.* when $(\bar{x}_{n'} \ge \bar{x}_n)$ more so when $(\bar{x}_{n'}/\bar{x}_n) \ge 1.2$. Also for the proposed Bootstrap methods the Bootstrap and its linearised version seem to perform very similarly as is the case with the Jackknife. Note here that v_{B_2} and the linearised v_B were not distinguishable. Only for balanced samples, for which $(\bar{x}_{n'}/\bar{x}_n) \cong 1$, do the usual variance estimator v_0 and v_1 perform well. It is emphasized here that the comparisons have been done with respect to the conditional mean squared error MSE_c (represented by the thick line in Fig 1) and the group MSE represented on the vertical axis in Fig 1 should not be misinterpreted to imply that lower the MSE, the better the performance. Rather, if some parts of the graph are above the conditional MSE then it is over-estimating the true MSE at such points and if they are below it, then under-estimating. The graphs of the proposed Bootstrap variance estimators coincide with that of Rao and Sitter's Jackknife at almost all points, supporting that the proposed Bootstrap methods are almost as efficient as the Jackknife method.

REFERENCES

[1] Biemer, P.P. and Atkinson, D. (1993). Estimation of measurement bias using a model prediction approach. *Survey Methodology*, **19**, 127-136.

[2] Fuller, W.A. (1998). Replication variance estimation for two-phase samples. *Statist. Sinica*, **8**, 1153-1164.

[3] Kott, P.S. and Stukel, D.M. (1997). Can the Jackknife be used with a two-phase sample? *Survey Methodology*, **23**, 81-89.

[4] Rao, J.N.K. (1985). Conditional inference in survey sampling. *Survey Methodology*, **11**, 15-31.

[5] Rao, J.N.K. and Sitter, R.R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika*, **82**, 453-460.

[6] Rao, J.N.K. and Sitter, R.R. (1997). Variance estimation under stratified two-phase sampling with applications to measurement bias. In : Survey Measurement and Process Quality, Wiley Series in Probability and Statistics (ed. L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz and D. Trewin), John Wiley and Sons Inc., New York, 753-768.

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

[8] Schreuder, H.T., Li, H.G. and Scott, C.T. (1987). Jackknife and Bootstrap estimation for sampling with partial replacement. *Forest Science*, **33**, 676-689.

[9] Sitter, R.R. (1997). Variance estimation for the regression estimator in two-phase sampling. *Jour. Amer. Statist. Assoc.*, **92**, 780 – 787.

[10] Stephan, F.F. (1945). The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate. *Ann. Math. Stat.*, **16**, 50-61.