Pak. J. Statist.
2015 Vol. 31(4), 399-416

VARIANCE ESTIMATION IN TWO-PHASE SAMPLING

Raghunath Arnab[1,4§], **Sarjinder Singh**[2], **Antonio Arcos**[3]
and **Kago Kebotsamang**[4]

[1] University of KwaZulu-Natal, RSA
Email: arnabr@mopipi.ub.bw
[2] Texas A& M University, Kingsville, USA
[3] Universidad de Granada, Spain
[4] University of Botswana, Botswana
[§] Corresponding Author

ABSTRACT

The problem of variance estimation for two-phase sampling was considered by Särndal et al. (1992). Hidiroglou et al. (2009) studied the properties of the proposed estimators theoretically while Haziza et al. (2001) empirically studied the performances of the proposed estimators. In this paper a few alternative variance estimators have been proposed. The proposed variance estimators are found to be more efficient than the existing estimators when the study variable is proportional to the auxiliary variable. Empirical investigations also reveal that the proposed variance estimators fare better in respect of efficiency and preserving non-negativity property.

KEYWORDS

Two-phase sampling, variance estimation, varying probability sampling.

1. INTRODUCTION

The two-phase sampling is used when the variable under study y is well related to the auxiliary variable x but the information of the auxiliary variable is not available before the survey and the cost of collecting data for the auxiliary variable x is much cheaper than the study variable. Hence, in two-phase sampling, a relative large sample is initially selected from the population and only information of the auxiliary variable x is collected. Then in phase two another sample is selected either from the sample selected in the first-phase or from the entire population independently and information of the study variable y is obtained. The two-phase sampling was introduced by Neyman (1938). Recently, Särndal et al. (1992) proposed the double expansion estimator for estimating the population total when the sample selected on both phases by arbitrary sampling design. They also proposed Horvitz-Thopmson (HT, 1952) and Yates-Grundy (YG, 1953) type variance estimators of the proposed double expansion (DE) estimator. Hidiroglou et al. (2009) pointed out that the proposed HT type variance estimator can be very unstable and can even take negative values. If the variance estimator becomes negative it cannot be used for measuring the magnitude of the sampling error, interval estimation and testing of hypotheses among others. Hidiroglou et al. (2009) also studied the properties of the HT and YG type variance estimators theoretically in details. Haziza et al. (2011) considered DE-estimator and Hajèk (1971) estimator of the population total for the two-phase sampling. They also proposed HT and YG type variance estimators for Hajèk (HA) estimator. Haziza et al. (2011) studied the properties of their variance estimators in terms of stability and non-negativity theoretically as well as through extensive simulation studies. They found that HT-type variance estimators are more prone to take negative values than YG type variance estimators.

In this present paper we have proposed few alternative variance estimators which are modifications of the variance estimators investigated by Haziza et al. (2011). The proposed estimators are found better in terms of efficiency and are less prone to take negative values.

1.1 HT and YG Variance estimators in two-phase sampling

Consider a finite population $U = \{1, ..., i, ..., N\}$ of N identifiable units. In the first phase, a sample s_1 of size n_1 units is selected using a sampling design p_1 and information of the auxiliary variable x is observed. Let π_{1i} and $\pi_{1ij}(>0)$ be the inclusion probabilities for the i th, and i th and j th unit ($i \neq j$) respectively. In the second phase, a sub-sample s_2 of size n_2 is selected from s_1 by following a sampling design $p_{2|s_1}$ with inclusion probabilities $\pi_{2i|s_1}$ and $\pi_{2ij|s_1}(>0)$ for the i th, and i th and j th unit ($i \neq j$) respectively.

Our objective is to estimate the finite population total $Y = \sum_{i \in U} y_i$ using a two-phase sample $s = (s_1, s_2)$ selected by some suitable sampling designs.

Särndal et al. (1992) proposed the following Double expansion (DE) estimator for the population total Y :

$$\hat{Y}_{DE} = \sum_{i \in s_2} \frac{y_i}{\pi^*_{i|s_1}} \tag{1.1}$$

where $\pi^*_{i|s_1} = \pi_{1i}\pi_{2i|s_1}$.

Haziza et al. (2011) proposed HA type estimator of Y as follows:

$$\hat{Y}_{HA} = \frac{N}{\hat{N}_{DE}} \hat{Y}_{DE} \tag{1.2}$$

where $\hat{N}_{DE} = \sum_{i \in s_2} \frac{1}{\pi^*_{i|s_1}}$.

The estimator $\hat{Y}_{DE}$ is design-unbiased for Y while $\hat{N}_{DE}$ is asymptotically design unbiased.

1.2 Double Expansion Estimator $\hat{Y}_{DE}$

The variance of $\hat{Y}_{DE}$ is given by

$$\begin{aligned} V_{DE} &= V\left(\hat{Y}_{DE}\right) = V\left(E\left(\hat{Y}_{DE} \mid s_1\right)\right) + E\left(V\left(\hat{Y}_{DE} \mid s_1\right)\right) \\ &= V\left(\sum_{i\in s_1} \frac{y_i}{\pi_{1i}}\right) + E\left[V\left(\sum_{i\in s_2} \frac{y_i}{\pi^*_{i|s_1}} \mid s_1\right)\right] \\ &= Q_1(y) + Q_2(y) \end{aligned} \tag{1.3}$$

where $Q_1(y) = V\left(\sum_{i\in s_1} \frac{y_i}{\pi_{1i}}\right)$ and $Q_2(y) = E\left(V\left(\hat{Y}_{DE} \mid s_1\right)\right)$.

Now

$$\begin{aligned} Q_1(y) &= \sum_{i\in U}\left(\frac{1}{\pi_{1i}} - 1\right) y_i^2 + \sum_{i\neq}\sum_{j\in U}\left(\frac{\pi_{1ij}}{\pi_{1i}\pi_{1j}} - 1\right) y_i y_j \\ &= Q_1^{ht}(y) \end{aligned}$$

For fixed sample size design $Q_1(y)$ can be written as

$$\begin{aligned} Q_1(y) &= \frac{1}{2}\sum_{i\neq}\sum_{j\in U}\left(\pi_{1i}\pi_{1j} - \pi_{1ij}\right)\left(\frac{y_i}{\pi_{1i}} - \frac{y_j}{\pi_{1j}}\right)^2 \\ &= Q_1^{yg}(y) \end{aligned}$$

From the expressions $Q_1^{ht}(y)$ and $Q_1^{yg}(y)$, HT-type and YG-type unbiased estimators of $Q_1(y)$ are obtained as follows:

$$\hat{Q}_1^{ht}(y) = \sum_{i\in s_2} \frac{(1-\pi_{1i})}{\pi_{2i|s_1}}\left(\frac{y_i}{\pi_{1i}}\right)^2 - \sum_{i\neq}\sum_{j\in s_2} \frac{\Delta_{1ij}}{\pi^*_{ij|s_1}} \frac{y_i}{\pi_{1i}} \frac{y_j}{\pi_{1j}}$$

and

$$\hat{Q}_1^{yg}(y) = \frac{1}{2}\sum_{i\neq}\sum_{j\in s_2} \frac{\Delta_{1ij}}{\pi^*_{ij|s_1}}\left(\frac{y_i}{\pi_{1i}} - \frac{y_j}{\pi_{1j}}\right)^2 \tag{1.4}$$

where $\Delta_{1ij} = \pi_{1i}\pi_{1j} - \pi_{1ij}$ and $\pi^*_{ij|s_1} = \pi_{2ij|s_1}\pi_{1ij}$.

Similarly, writing

$$V\left(\hat{Y}_{DE} \mid s_1\right)=\sum_{i\in s_1}\left(\frac{1}{\pi_{2i|s_1}}-1\right)\left(\frac{y_i}{\pi_{1i}}\right)^2+\sum_{i\neq}\sum_{j\in s_1}\left(\frac{\pi_{2ij|s_1}}{\pi_{2i|s_1}\pi_{2j|s_1}}-1\right)\frac{y_i}{\pi_{1i}}\frac{y_j}{\pi_{1j}}=Q_{2|s_1}^{ht}(y)$$

$$=\frac{1}{2}\sum_{i\neq}\sum_{j\in s_1}\left(\pi_{2i|s_1}\pi_{2j|s_1}-\pi_{2ij|s_1}\right)\left(\frac{y_i}{\pi_{1i}\pi_{2i|s_1}}-\frac{y_j}{\pi_{1j}\pi_{2j|s_1}}\right)^2=Q_{2|s_1}^{yg}(y),$$

two unbiased estimators HT-type and YG type of $Q_2(y)$ are obtained as

$$\hat{Q}_2^{ht}(y)=\sum_{i\in s_2}\left(1-\pi_{2i|s_1}\right)\left(\frac{y_i}{\pi_{1i}\pi_{2i|s_1}}\right)^2-\sum_{i\neq}\sum_{j\in s_2}\frac{\Delta_{2ij|s_1}}{\pi_{2ij|s_1}}\frac{y_i}{\pi_{1i}\pi_{2i|s_1}}\frac{y_j}{\pi_{1j}\pi_{2j|s_1}}$$

$$=\sum_{i\in s_2}\left(1-\pi_{2i|s_1}\right)\left(\frac{y_i}{\pi^*_{i|s_1}}\right)^2-\sum_{i\neq}\sum_{j\in s_2}\frac{\Delta_{2ij|s_1}}{\pi_{2ij|s_1}}\frac{y_i}{\pi^*_{i|s_1}}\frac{y_j}{\pi^*_{j|s_1}} \tag{1.5}$$

and

$$\hat{Q}_2^{yg}=\frac{1}{2}\sum_{i\neq}\sum_{j\in s_2}\frac{\Delta_{2ij|s_1}}{\pi_{2ij|s_1}}\left(\frac{y_i}{\pi_{1i}\pi_{2i|s_1}}-\frac{y_j}{\pi_{1j}\pi_{2j|s_1}}\right)^2$$

$$=\frac{1}{2}\sum_{i\neq}\sum_{j\in s_2}\frac{\Delta_{2ij|s_1}}{\pi_{2ij|s_1}}\left(\frac{y_i}{\pi^*_{i|s_1}}-\frac{y_j}{\pi^*_{j|s_1}}\right)^2 \tag{1.6}$$

where $\pi^*_{i|s_1}=\pi_{1i}\pi_{2i|s_1}$ and $\Delta_{2ij|s_1}=\left(\pi_{2i|s_1}\pi_{2j|s_1}-\pi_{2ij|s_1}\right)$.

From the expressions (1.3) to (1.6), unbiased estimators of the variance of the total $\hat{Y}_{DE}$ are obtained as follows:

$$\hat{V}_{DE}^{ht}=\hat{Q}_1^{ht}(y)+\hat{Q}_2^{ht}(y) \tag{1.7}$$

and

$$\hat{V}_{DE}^{yg}=\hat{Q}_1^{yg}(y)+\hat{Q}_2^{yg}(y) \tag{1.8}$$

The HT-type and YG-type variance estimators $\hat{V}_{DE}^{ht}$ and $\hat{V}_{DE}^{yg}$ were proposed respectively by Särndal et al. (1992) and Hidiroglou et al. (2009) respectively.

It is well known that both the estimators $\hat{V}_{DE}^{ht}$ and $\hat{V}_{DE}^{yg}$ can take negative values. The estimator $\hat{V}_{DE}^{yg}$ is always non-negative for fixed sample size designs for which Δ_{1ij} and $\Delta_{2ij|s_1}$ are both nonnegative. We will call such sampling designs as $P*$. However for such $P*$ designs, $\hat{V}_{DE}^{ht}$ can take negative values.

Hidiroglou et al. (2009) and Haziza et al. (2011) studied in detail the properties of stability and non-negativity of the variance estimators $\hat{V}_{DE}^{ht}$ and $\hat{V}_{DE}^{yg}$. Haziza et al. (2011)

pointed out that the estimators $\hat{Q}_1^{yg}(y)=0$ if $CV(y/\pi_1)=0$ and $\hat{Q}_2^{yg}(y)=0$ if $CV(y/\pi^*)=0$ equal to zero. On the other hand, the components $\hat{Q}_1^{ht}(y)$ or $\hat{Q}_2^{ht}(y)$ becomes negative to satisfy unbiasedness condition of $\hat{V}_{DE}^{ht}$. So, in this situation $\hat{V}_{DE}^{ht}$ can take negative values. Hence if $CV(y/\pi_1)$ and / or $CV(y/\pi^*)$ is small, $\hat{V}_{DE}^{yg}$ is more stable than $\hat{V}_{DE}^{ht}$. However if $CV(y/\pi_1)$ and / or $CV(y/\pi^*)$ is large, we cannot compare stability of the variance estimators.

1.3 Hajèk Estimator $\hat{Y}_{HA}$

The Hajèk estimator $\hat{Y}_{HA}$ is a ratio estimator and hence to find the variance of $\hat{Y}_{HA}$, Haziza et al. (2011) considered the following approximation:

$$\begin{aligned}\hat{Y}_{HA}-Y &= \frac{N}{\sum_{i\in s_2}\frac{1}{\pi^*_{i|s_1}}}\left(\sum_{i\in s_2}\frac{y_i}{\pi^*_{i|s_1}}-\bar{Y}\sum_{i\in s_2}\frac{1}{\pi^*_{i|s_1}}\right)\\ &\cong\left(\sum_{i\in s_2}\frac{y_i-\bar{Y}}{\pi^*_{i|s_1}}\right)\\ &=\sum_{i\in s_2}\frac{z_i}{\pi^*_{i|s_1}} \qquad (1.9)\end{aligned}$$

where $z_i=y_i-\bar{Y}$ and $\bar{Y}=Y/N$.

Under the approximation (1.9), the variance of $\hat{Y}_{HA}$ was obtained as

$$\begin{aligned}V_{HA}=V\left(\hat{Y}_{HA}\right)&=V\left(\sum_{i\in s_1}\frac{z_i}{\pi_{1i}}\right)+E\left(V\left(\sum_{i\in s_2}\frac{z_i}{\pi^*_{i|s_1}}\mid s_1\right)\right)\\ &=Q_1(z)+Q_2(z) \qquad (1.10)\end{aligned}$$

where $Q_1(z)$ and $Q_2(z)$ are obtained by substituting $y_i=z_i$ in the expressions of $Q_1(y)$ and $Q_2(y)$ respectively.

Haziza et al. (2011) proposed the following variance estimators analogous to $\hat{V}_{DE}^{ht}$ and $\hat{V}_{DE}^{yg}$:

$$\hat{V}_{HA}^{ht}=\hat{Q}_1^{ht}(\hat{z})+\hat{Q}_2^{ht}(\hat{z}) \qquad (1.11)$$

and

$$\hat{V}_{HA}^{yg}=\hat{Q}_1^{yg}(\hat{z})+\hat{Q}_2^{yg}(\hat{z}) \qquad (1.12)$$

where $\hat{Q}_i^{j}(\hat{z})$ was obtained from $\hat{Q}_t^{j}(y)$ by replacing y_i with

$$\hat{z}_i = \frac{N}{\sum\limits_{i\in s_2} \frac{1}{\pi^*_{i|s_1}}} \left(y_i - \frac{\sum\limits_{i\in s_2} \frac{y_i}{\pi^*_{i|s_1}}}{\sum\limits_{i\in s_2} \frac{1}{\pi^*_{i|s_1}}} \right) \text{for } t = 1,2 \text{ and } j = ht, yg \tag{1.13}$$

Here also $\hat{V}_{HA}^{yg}$ estimator is always nonnegative for $P*$ sampling designs. Haziza et al. (2011) studied the properties of the estimators $\hat{V}_{HA}^{ht}$ and $\hat{V}_{HA}^{yg}$ but no straight forward conclusion relating to non-negativity and stability properties of the estimators were obtained.

2. PROPOSED VARIANCE ESTIMATORS

Haziza et al. (2011) use the auxiliary variable x collected in the phase one as a measure of size for the selection of sample in phase two. Since the variable x is expected to be highly related to the study variable y, we can improve the above mentioned variance estimators by extracting relationship between x and y at the stages of estimation also.

The proposed improved variance estimators for $\hat{Y}_{DE}$ are as follows:

$$\hat{\hat{V}}_{DE}^{ht} = \hat{\hat{Q}}_1^{ht}(y,x) + \hat{\hat{Q}}_2^{ht}(y,x) \tag{2.1}$$

and

$$\hat{\hat{V}}_{DE}^{yg} = \hat{\hat{Q}}_1^{yg}(y,x) + \hat{\hat{Q}}_2^{yg}(y,x) \tag{2.2}$$

where

$$\hat{\hat{Q}}_1^{ht}(y,x) = \frac{\hat{Q}_1^{ht}(y)}{\hat{Q}_1^{ht}(x)} E\left(\hat{Q}_1^{ht}(x) \mid s_1\right),\ \hat{\hat{Q}}_2^{ht}(y,x) = \frac{\hat{Q}_2^{ht}(y)}{\hat{Q}_2^{ht}(x)} E\left(\hat{Q}_2^{ht}(x) \mid s_1\right),$$

$$\hat{\hat{Q}}_1^{yg}(y,x) = \frac{\hat{Q}_1^{yg}(y)}{\hat{Q}_1^{yg}(x)} E\left(\hat{Q}_1^{yg}(x) \mid s_1\right) \text{ and } \hat{\hat{Q}}_2^{yg}(y,x) = \frac{\hat{Q}_2^{yg}(y)}{\hat{Q}_2^{yg}(x)} E\left(\hat{Q}_2^{yg}(x) \mid s_1\right);\ \hat{Q}_r^t(x)$$

is obtained from $\hat{Q}_r^t(y)$ by replacing $y_i = x_i$ for $r = 1,2$ and $t = ht, yg$;

$$E\left(\hat{Q}_1^{ht}(x) \mid s_1\right) = E\left[\sum_{i\in s_2} \frac{(1-\pi_{1i})}{\pi_{2i|s_1}} \left(\frac{x_i}{\pi_{1i}}\right)^2 - \sum_{i\neq} \sum_{j\in s_2} \frac{\Delta_{1ij}}{\pi^*_{ij|s_1}} \frac{x_i}{\pi_{1i}} \frac{x_j}{\pi_{1j}}\right]$$
$$= \left[\sum_{i\in s_1} (1-\pi_{1i}) \left(\frac{x_i}{\pi_{1i}}\right)^2 - \sum_{i\neq} \sum_{j\in s_1} \frac{\Delta_{1ij}}{\pi_{1ij}} \frac{x_i}{\pi_{1i}} \frac{x_j}{\pi_{1j}}\right]$$

$$E\left[\hat{Q}_2^{ht}(x)\mid s_1\right]=E\left[\sum_{i\in s_2}\left(1-\pi_{2i|s_1}\right)\left(\frac{x_i}{\pi^*_{i|s_1}}\right)^2-\sum_{i\neq}\sum_{j\in s_2}\frac{\Delta_{2ij|s_1}}{\pi_{2ij|s_1}}\frac{x_i}{\pi^*_{i|s_1}}\frac{x_j}{\pi^*_{j|s_1}}\right]$$

$$=\left[\sum_{i\in s_1}\left(1-\pi_{2i|s_1}\right)\pi_{2i|s_1}\left(\frac{x_i}{\pi^*_{i|s_1}}\right)^2-\sum_{i\neq}\sum_{j\in s_1}\Delta_{2ij|s_1}\frac{x_i}{\pi^*_{i|s_1}}\frac{x_j}{\pi^*_{j|s_1}}\right]$$

$$E\left(\hat{Q}_1^{yg}(x)\mid s_1\right)=\frac{1}{2}E\left[\sum_{i\neq}\sum_{j\in s_2}\frac{\Delta_{1ij}}{\pi^*_{ij|s_1}}\left(\frac{x_i}{\pi_{1i}}-\frac{x_j}{\pi_{1j}}\right)^2\right]$$

$$=\frac{1}{2}\left[\sum_{i\neq}\sum_{j\in s_1}\frac{\Delta_{1ij}}{\pi_{1ij}}\left(\frac{x_i}{\pi_{1i}}-\frac{x_j}{\pi_{1j}}\right)^2\right]$$

$$E\left(\hat{Q}_2^{yg}(x)\mid s_1\right)=\frac{1}{2}E\left[\sum_{i\neq}\sum_{j\in s_2}\frac{\Delta_{2ij|s_1}}{\pi_{2ij|s_1}}\left(\frac{x_i}{\pi^*_{i|s_1}}-\frac{x_j}{\pi^*_{j|s_1}}\right)^2\right]$$

$$=\frac{1}{2}\left[\sum_{i\neq}\sum_{j\in s_1}\Delta_{2ij|s_1}\left(\frac{x_i}{\pi^*_{i|s_1}}-\frac{x_j}{\pi^*_{j|s_1}}\right)^2\right]$$

Similarly, the proposed two alternative variance estimators for $\hat{Y}_{HA}$ are given as follows:

$$\hat{\hat{V}}_{HA}^{ht}=\hat{\hat{Q}}_1^{ht}\left(\hat{z},\hat{x}\right)+\hat{\hat{Q}}_2^{ht}\left(\hat{z},\hat{x}\right) \tag{2.3}$$

and

$$\hat{\hat{V}}_{HA}^{yg}=\hat{\hat{Q}}_1^{yg}\left(\hat{z},\hat{x}\right)+\hat{\hat{Q}}_2^{yg}\left(\hat{z},\hat{x}\right) \tag{2.4}$$

where

$$\hat{\hat{Q}}_1^{ht}(\hat{z},\hat{x})=\frac{\hat{Q}_1^{ht}(\hat{z})}{\hat{Q}_1^{ht}(\hat{x})}E\left(\hat{Q}_1^{ht}(\hat{x})\mid s_1\right),\ \hat{\hat{Q}}_2^{ht}(\hat{z},\hat{x})=\frac{\hat{Q}_2^{ht}(\hat{z})}{\hat{Q}_2^{ht}(\hat{x})}E\left(\hat{Q}_2^{ht}(\hat{x})\mid s_1\right),$$

$$\hat{\hat{Q}}_1^{yg}(\hat{z},\hat{x})=\frac{\hat{Q}_1^{yg}(\hat{z})}{\hat{Q}_1^{yg}(\hat{x})}E\left(\hat{Q}_1^{yg}(\hat{x})\mid s_1\right)\text{ and }\hat{\hat{Q}}_2^{yg}(\hat{z},\hat{x})=\frac{\hat{Q}_2^{yg}(\hat{z})}{\hat{Q}_2^{yg}(\hat{x})}E\left(\hat{Q}_2^{yg}(\hat{x})\mid s_1\right);\hat{Q}_r^t(\hat{x})$$

is obtained from $\hat{Q}_r^t(y)$ by replacing

$$y_i=\hat{x}_i=\frac{N}{\sum_{i\in s_2}\frac{1}{\pi^*_{i|s_1}}}\left(x_i-\frac{\sum_{i\in s_2}\frac{x_i}{\pi^*_{i|s_1}}}{\sum_{i\in s_2}\frac{1}{\pi^*_{i|s_1}}}\right)\text{for } r=1,2 \text{ and } t=ht,yg \tag{2.5}$$

$$E\left(\hat{Q}_1^{ht}(\hat{\hat{x}})\mid s_1\right)=\left[\sum_{i\in s_1}\left(1-\pi_{1i}\right)\left(\frac{\hat{\hat{x}}_i}{\pi_{1i}}\right)^2-\sum_{i\neq}\sum_{j\in s_1}\frac{\Delta_{1ij}}{\pi_{1ij}}\frac{\hat{\hat{x}}_i}{\pi_{1i}}\frac{\hat{\hat{x}}_j}{\pi_{1j}}\right],$$

$$E\left[\hat{Q}_2^{ht}\left(\hat{\hat{x}}\right)\mid s_1\right]=\left[\sum_{i\in s_1}\left(1-\pi_{2i|s_1}\right)\pi_{2i|s_1}\left(\frac{\hat{\hat{x}}_i}{\pi^*_{i|s_1}}\right)^2-\sum_{i\neq}\sum_{j\in s_1}\Delta_{2ij|s_1}\frac{\hat{\hat{x}}_i}{\pi^*_{i|s_1}}\frac{\hat{\hat{x}}_j}{\pi^*_{j|s_1}}\right],$$

$$E\left(\hat{Q}_1^{yg}\left(\hat{\hat{x}}\right)\mid s_1\right)=\frac{1}{2}\left[\sum_{i\neq}\sum_{j\in s_1}\frac{\Delta_{1ij}}{\pi_{1ij}}\left(\frac{\hat{\hat{x}}_i}{\pi_{1i}}-\frac{\hat{\hat{x}}_j}{\pi_{1j}}\right)^2\right]$$

and

$$E\left(\hat{Q}_2^{yg}\left(\hat{\hat{x}}\right)\mid s_1\right)=\frac{1}{2}\left[\sum_{i\neq}\sum_{j\in s_1}\Delta_{2ij|s_1}\left(\frac{\hat{\hat{x}}_i}{\pi^*_{i|s_1}}-\frac{\hat{\hat{x}}_j}{\pi^*_{j|s_1}}\right)^2\right]$$

where

$$\hat{\hat{x}}_i=\frac{N}{\sum_{i\in s_1}\frac{1}{\pi_{1i}}}\left(x_i-\frac{\sum_{i\in s_1}\frac{x_i}{\pi_{1i}}}{\sum_{i\in s_1}\frac{1}{\pi_{1i}}}\right). \tag{2.6}$$

Example 2.1.

Consider the situation where the first phase sample s_1 is selected by SRSWOR method and the second-phase sample s_2 by arbitrary sampling designs. In this case, we have

$$\pi_{1i}=n_1/N\,,\ \pi_{1ij}=\frac{n_1(n_1-1)}{N(N-1)}\,,\ \hat{Y}_{DE}=\frac{N}{n_1}\sum_{i\in s_2}\frac{y_i}{\pi_{2i|s_1}}\,,$$

$$\hat{Q}_1^{ht}(y)=\frac{N-n_1}{n_1}\left[\sum_{i\in s_2}\frac{y_i^2}{\pi_{2i|s_1}}-\frac{1}{N-1}\sum_{i\neq}\sum_{j\in s_2}\frac{y_iy_j}{\pi_{2ij|s_1}}\right]$$

$$\hat{Q}_2^{ht}(y)=\left(\frac{n_1}{N}\right)^2\left[\sum_{i\in s_2}\left(1-\pi_{2i|s_1}\right)\left(\frac{y_i}{\pi_{2i|s_1}}\right)^2-\sum_{i\neq}\sum_{j\in s_2}\frac{\Delta_{2ij|s_1}}{\pi_{2ij|s_1}}\frac{y_i}{\pi_{2i|s_1}}\frac{y_j}{\pi_{2j|s_1}}\right],$$

$$\hat{Q}_1^{yg}(y)=\frac{1}{2}\frac{N\left(N-n_1\right)}{n_1^2\left(n_1-1\right)}\sum_{i\neq}\sum_{j\in s_2}\frac{\left(y_i-y_j\right)^2}{\pi_{2ij|s_1}}\,,$$

$$\hat{Q}_2^{yg}(y) = \frac{1}{2}\left(\frac{n_1}{N}\right)^2 \sum_{i\neq} \sum_{j\in s_2} \frac{\Delta_{2ij|s_1}}{\pi_{2ij|s_1}} \left(\frac{y_i}{\pi_{2i|s_1}} - \frac{y_j}{\pi_{2j|s_1}}\right)^2,$$

$$E\left(\hat{Q}_1^{ht}(x) \mid s_1\right) = \frac{N(N-n_1)}{n_1(n_1-1)} \sum_{i\in s_1} \left(x_i - \bar{x}_{s_1}\right)^2, \ \bar{x}_{s_1} = \sum_{i\in s_1} x_i / n_1,$$

$$E\left[\hat{Q}_2^{ht}(x) \mid s_1\right] = \left(\frac{N}{n_1}\right)^2 \left[\sum_{i\in s_1} \left(1-\pi_{2i|s_1}\right)\pi_{2i|s_1} \left(\frac{x_i}{\pi_{2i|s_1}}\right)^2 - \sum_{i\neq} \sum_{j\in s_1} \Delta_{2ij|s_1} \frac{x_i}{\pi_{2i|s_1}} \frac{x_j}{\pi_{2j|s_1}}\right],$$

$$E\left(\hat{Q}_1^{yg}(x) \mid s_1\right) = \frac{N(N-n_1)}{n_1^2(n_1-1)} \sum_{i\neq} \sum_{j\in s_1} \left(x_i - x_j\right)^2,$$

$$E\left(\hat{Q}_2^{yg}(x)\right) = \frac{1}{2}\left(\frac{N}{n_1}\right)^2 \sum_{i\neq} \sum_{j\in s_1} \Delta_{2ij|s_1} \left(\frac{x_i}{\pi_{2i|s_1}} - \frac{x_j}{\pi_{2j|s_1}}\right)^2$$

Example 2.2

Consider the situation where the first phase sample s_1 is selected by Lahiri-Midzuno-Sen (1951, 1952, 1953) sampling scheme and the 2nd phase sample s_2 is selected by arbitrary sampling scheme. In Lahiri-Midzuno-Sen (LMS) sampling scheme, on the first draw one unit is selected with normed size measure $p_i\left(p_i > 0, \sum_i p_i = 1\right)$ attached to the i th unit and the remaining $n_1 - 1$ are selected from those $N-1$ units which were not selected in the first draw by simple random sampling without replacement (SRSWOR) method. The first and second order inclusion probabilities for the LMS sampling design are given by $\pi_{1i} = \frac{N-n_1}{N-1} p_i + \frac{n_1-1}{N-1}$ and $\pi_{1ij} = \left(p_i + p_j\right)\frac{(N-n_1)(n_1-1)}{(N-1)(N-2)} + \frac{(n_1-1)(n_1-2)}{(N-1)(N-2)}$ respectively. The expressions of $\hat{Q}_j^{\theta}$ and $E\left(\hat{Q}_j^{\theta}(x) \mid s_1\right)$, $j = 1,2; \theta = ht, yg$ are highly complex and they are omitted here.

2.1 Properties of the Proposed Variance Estimators

Since the estimators $\hat{\hat{Q}}_1^{ht}(y,x)$, $\hat{\hat{Q}}_2^{ht}(y,x)$, $\hat{\hat{Q}}_1^{yg}(y,x)$ and $\hat{\hat{Q}}_2^{yg}(y,x)$ are ratio type estimators, the proposed variance estimator $\hat{\hat{V}}_{DE}^{ht}$ and $\hat{\hat{V}}_{DE}^{yg}$ are biased estimator for V_{DE}. Similarly, $\hat{\hat{V}}_{HA}^{ht}$ and $\hat{\hat{V}}_{HA}^{yg}$ are biased estimator for V_{HA}. The amount of bias is expected to be less if the sample size at the second phase is large. The variance estimators $\hat{\hat{V}}_{DE}^{yg}$ and $\hat{\hat{V}}_{HA}^{yg}$ are always nonnegative if $P*$ design is used at both the phases of sampling. In case the study y variable is approximately proportional to the auxiliary variable x, $\hat{Q}_1^{ht}(y)$ and $\hat{Q}_1^{ht}(x)$ are expected to have the same sign and the sign of $\hat{\hat{Q}}_1^{ht}(y,x)$ will be determined by the sign of $E\left(\hat{Q}_1^{ht}(x)\,|\,s_1\right)$. Since the quantity $E\left(\hat{Q}_1^{ht}(x)\,|\,s_1\right)$ is based on the larger sample of size n_1, it should have higher concentration around its expected value $Q_1(x)$ (which is positive) than $\hat{Q}_1^{ht}(y)$. Hence $E\left(\hat{Q}_1^{ht}(x)\,|\,s_1\right)$ is expected to take more positive values than $\hat{Q}_1^{ht}(y)$ and $\hat{Q}_1^{ht}(x)$. Thus the proposed estimators $\hat{\hat{V}}_{DE}^{ht}$ and $\hat{\hat{V}}_{HA}^{ht}$ are likely to take less negative values than $\hat{V}_{DE}^{ht}$ and $\hat{V}_{HA}^{ht}$.

As efficiency is concerned, we show that all the proposed variance estimators $\hat{\hat{V}}_{DE}^{ht}, \hat{\hat{V}}_{DE}^{yg}, \hat{\hat{V}}_{HA}^{ht}$ and $\hat{\hat{V}}_{HA}^{yg}$ are more efficient than those considered by Haziza et al. (2011) if the study variable y is proportional to the auxiliary variable x.

Theorem 2.1

If the study variable y is proportional to the auxiliary variable x, then

$$Var\left(\hat{\hat{V}}_j^{\alpha}\right) \le Var\left(\hat{V}_j^{\alpha}\right);\ \ j = DE, HA \text{ and } \alpha = ht, yg$$

Proof:

Let

$$y = \beta x \text{, then } \hat{\hat{Q}}_1^{ht}(y,x) = \beta^2 E\left(\hat{Q}_1^{ht}(x)\,|\,s_1\right)$$
$$= E\left(\hat{Q}_1^{ht}(y)\,|\,s_1\right)$$

Now

$$\begin{aligned}
Var\left(\hat{V}_{DE}^{ht}\right) &= Var\left(\hat{Q}_1^{ht}(y) + \hat{Q}_2^{ht}(y)\right) \\
&= E\left\{Var\left(\hat{Q}_1^{ht}(y) + \hat{Q}_2^{ht}(y)\,|\,s_1\right)\right\} + Var\left\{E\left(\hat{Q}_1^{ht}(y) + \hat{Q}_2^{ht}(y)\,|\,s_1\right)\right\} \\
&\ge Var\left\{E\left(\hat{Q}_1^{ht}(y) + \hat{Q}_2^{ht}(y)\,|\,s_1\right)\right\} \\
&= Var\left(\hat{\hat{Q}}_1^{ht}(y,x) + \hat{\hat{Q}}_2^{ht}(y,x)\right) \\
&= Var\left(\hat{\hat{V}}_{DE}^{ht}\right)
\end{aligned}$$

Similarly we can show $Var\left(\hat{\hat{V}}_{DE}^{yg}\right) \le Var\left(\hat{V}_{DE}^{ht}\right)$, $Var\left(\hat{\hat{V}}_{HA}^{ht}\right) \le Var\left(\hat{V}_{HA}^{ht}\right)$ and $Var\left(\hat{\hat{V}}_{HA}^{yg}\right) \le Var\left(\hat{\hat{V}}_{HA}^{ht}\right)$.

3. SIMULATION STUDIES

In this section, we investigated the performances of the proposed variance estimators with the existing estimators investigated by Haziza et al. (2011). For the simulation studies, we generated 8 populations of which 4 were of size 100 and the remaining were of size 200 each. Each population consisted of two variables: the variable of interest y and auxiliary variable x as was considered by Haziza et al. (2011). The size variable x was generated from a gamma distribution with the scale parameter $\gamma = 4$ and shape parameter $\theta = 25$. For given values of x, the values of y were generated using the model $y_i = \alpha + 2x_i + \in_i$ where $\in_i$ was assumed to be normally distributed with mean zero and variance $\sigma^2 x_i^g$. The 8 populations were generated by choosing $\alpha = 0.0, 5$; $\sigma^2 = 1$ and $g = 0.0, 0.5$. The CV of x for each of the population of size 100 was 0.2023 while for the other populations of size 200 was 0.2031. Summary statistics for the generated populations, in terms of CV, are reported in Table 3.0.

At the first phase, we selected a sample s_1 of size n_1 (=20, 30, 40) by SRSWOR method so that $\pi_{1i} = n_1 / N$ and $\pi_{1ij} = n_1(n_1 - 1)/\{N(N-1)\}$ as was selected by Haziza et al. (2011). From the selected sample s_1, only the values of the auxiliary variable x was observed. In the second phase, a sub-sample s_2 of size n_2 (=10, 15, 20) was selected from s_1 by Sampford (1967) sampling scheme using x as measure of size variable. The Sampford sampling scheme is as follows:

On the first draw, the ith unit is selected from s_1 with probability $p_i(1) = x_i \Big/ \sum_{i \in s_1} x_i$. Then the remaining $(n_2 - 1)$ units are drawn with replacement from the entire sample s_1 with probability proportional to $\lambda_i = p_i(1) \Big/ \{1 - n_1 p_i(1)\}$ i.e. the probability of selecting ith unit at kth draw is $p_i(k) = \lambda_i \Big/ \sum_{j \in s_1} \lambda_j$, $k = 2,..,n_2$. The selected units are accepted as a sample if all the n_2 units happened to be distinct, otherwise the entire selection is discarded and this process is repeated until a set of n_2 distinct units is obtained. Sampford (1967) has shown that the inclusion probability for the selection of ith unit is $\pi_{2i|s_1} = n_2 p_i(1)$ and $\Delta_{2ij|s_1} = \pi_{2i|s_1} \cdot \pi_{2j|s_1} - \pi_{2ij|s_1} \geq 0$. The expression for the second order inclusion probabilities is not simple. However, approximate expression of $\pi_{2ij|s_1}$ correct to $O\left(n_1^{-4}\right)$, derived by Asok and Sukhatme (1976) is given for $n_2 \geq 3$ as is follows:

$$\pi_{2ij|s_1} = n_2(n_2-1)p_i(1)p_j(1)\Bigg[1+\Big\{\big(p_i(1)+p_j(1)\big)-\sum_{j\in s_1} p_j^2(1)\Big\}$$
$$+\Big\{2\big(p_i^2(1)+p_j^2(1)\big)-2\sum_{j\in s_1} p_j^3(1)-(n_2-2)p_i(1)p_j(1)$$
$$+(n_2-3)\big(p_i(1)+p_j(1)\big)\sum_{j\in s_1} p_j^2(1)-(n_2-3)\Big(\sum_{j\in s_1} p_j^2(1)\Big)^2\Big\}-(n_2-2)p_i(1)p_j(1)$$
$$+(n_2-3)\big(p_i(1)+p_j(1)\big)\sum_{j\in s_1} p_j^2(1)-(n_2-3)\Big(\sum_{j\in s_1} p_j^2(1)\Big)^2\Bigg] \quad (3.1)$$

From the selected sample $s=(s_1,s_2)$, the estimators $\hat{Y}_{DE},\hat{Y}_{HA}$ of the population total Y, the variance estimators $v_1=\hat{V}_{DE}^{ht}, v_2=\hat{V}_{DE}^{yg}, v_1^*=\hat{V}_{HA}^{ht}, v_2^*=\hat{V}_{HA}^{yg}$ proposed by Haziza et al. (2011) and proposed in this article $v_3=\hat{\hat{V}}_{DE}^{ht}, v_4=\hat{\hat{V}}_{DE}^{yg}, v_3^*=\hat{\hat{V}}_{HA}^{ht}, v_4^*=\hat{\hat{V}}_{HA}^{yg}$ are computed. The selection of sample (s_1,s_2) was repeated R (= 100,000) times. The values of the estimators $\hat{Y}_{DE}$, $\hat{Y}_{HA}$, v_j and v_j^* were computed at the rth iteration and are denoted respectively by $\hat{Y}_{DE}(r)$, $\hat{Y}_{HA}(r)$, $v_j(r)$ and $v_j^*(r)$; $j=1,..,4$. The variances of $\hat{Y}_{DE}$ and $\hat{Y}_{HA}$ are respectively given by:

$$V_{DE}=\frac{1}{R}\sum_{r=1}^{R}\left(\hat{Y}_{DE}(r)-\frac{1}{R}\sum_{r=1}^{R}\hat{Y}_{DE}(r)\right)^2 \text{ and } V_{HA}=\frac{1}{R}\sum_{r=1}^{R}\left(\hat{Y}_{HA}(r)-\frac{1}{R}\sum_{r=1}^{R}\hat{Y}_{HA}(r)\right)^2 \quad (3.2)$$

The percentage relative biases and mean-square errors of the variance estimators v_j and v_j^* are respectively obtained as follows:

Bias of $v_j = B(j)=\dfrac{\frac{1}{R}\sum_{r=1}^{R}v_j(r)-V_{DE}}{V_{DE}}\times 100$; mean square error of $v_j =\ M(j)=\frac{1}{R}\sum_{r=1}^{R}\left(v_j(r)-V_{DE}\right)^2$

Bias of $v_j^* = B^*(j)=\dfrac{\frac{1}{R}\sum_{r=1}^{R}v_j^*(r)-V_{HA}}{V_{HA}}\times 100$; mean square error of $v_j^* =\ M^*(j)=\frac{1}{R}\sum_{r=1}^{R}\left(v_j^*(r)-V_{HA}\right)^2$

A problem of simulation studies with variance estimators is that the parameter to be estimated (the variance) is unknown. Then, one should select an estimate to measure the stability of the variance estimator. Note that we used the Monte Carlo Variance as a measure of stability of a variance estimator rather than the Monte Carlo MSE.

It was found that $v_1\left(v_1^*\right)$ is much inferior (with respect to bias and mean square error) to $v_j\left(v_j^*\right)$ for $j=2,3,4$. Therefore, the relative efficiencies of v_j and v_j^* were computed with respect to v_2 and v_2^* respectively. They are respectively denoted by

$$E_j=\left(v_2/v_j\right)\times 100 \text{ and } E_j^*=\left(v_2^*/v_j^*\right)\times 100\,;\; j=2,3,4 \tag{3.3}$$

The % of relative biases of the estimators are presented in Table 3.1. The estimator v_1 possesses a large bias for all the eight populations considered here. The estimators v_2, v_3, v_4 possess small and almost equal relative absolute bias in all situations while $v_1^*, v_2^*, v_3^*, v_4^*$ possess moderate bias of equal magnitudes. It is very difficult to choose any particular estimator from v_2, v_3, v_4 since none of them have smaller bias in all the situations. The relative bias of the estimator v_2 is negative which implies that it underestimates the variance while the rest estimators have positive bias which indicates overestimation of the variance.

As regards to efficiencies of the estimators v_3, v_4 bring enormous gain in efficiency over v_2. Similarly, v_3^*, v_4^* are much more efficient than v_2^*. It is very difficult to choose between v_3 and v_4 since none is better in all the situations. The efficiencies of the estimators v_3^* and v_4^* are virtually the same in all situations. The relative efficiencies of the estimators are given in table 3.2. Among the seven estimators $v_1, v_2, v_3, v_4, v_1^*, v_2^*, v_3^*$ only v_1 and v_3 were found to take negative values. The proposed estimator v_3 was found to take negative values much less than the existing estimator v_1. The number of negative values taken by the estimators v_1 and v_3 are presented in table 3.3.

Table 3.0
Coefficients of Variation for the Population

Coefficient of variation (CV)	*N*=100				*N*=200			
	$\alpha = 0.0$		$\alpha = 5.0$		$\alpha = 0.0$		$\alpha = 5.0$	
	g		g		g		g	
	0.0	0.5	0.0	0.5	0.0	0.5	0.0	0.5
$CV(x)$	0.2023	0.2023	0.2023	0.2023	0.2031	0.2031	0.2031	0.2031
$CV(y)$	0.2026	0.2113	0.1976	0.1988	0.2034	0.2050	0.1977	0.2046
$CV(y/x)$	0.0044	0.0540	0.0073	0.0502	0.0054	0.0503	0.0082	0.0504
$\lvert CV(z/x) \rvert$	5.1554	5.2851	5.1609	5.2844	5.1549	5.2717	5.1531	5.2837
ρ_{xy}	0.9997	0.9651	0.9996	0.9710	0.9996	0.9716	0.9997	0.9701

Table 3.1
Relative Percentage Bias

$$\left(v_1 = \hat{V}_{DE}^{ht}, v_2 = \hat{V}_{DE}^{yg}, v_1^* = \hat{V}_{HA}^{ht}, v_2^* = \hat{V}_{HA}^{yg}, v_3 = \hat{\hat{V}}_{DE}^{ht}, v_4 = \hat{\hat{V}}_{DE}^{yg}, v_3^* = \hat{\hat{V}}_{HA}^{ht}, v_4^* = \hat{\hat{V}}_{HA}^{yg}\right)$$

Estimators	*N*=100				*N*=200			
	$\alpha = 0.0$		$\alpha = 5.0$		$\alpha = 0.0$		$\alpha = 5.0$	
	g		g		g		g	
	0.0	0.5	0.0	0.5	0.0	0.5	0.0	0.5
	$n_1 = 20$, $n_2 = 10$							
v_1	150.64	127.90	160.94	136.73	132.53	106.55	141.83	114.12
v_2	-1.78	-0.86	-1.93	-1.09	-4.56	-3.15	-4.72	-3.45
v_3	1.91	-4.00	3.23	-3.16	0.54	-5.86	1.79	-5.53
v_4	2.24	4.73	2.31	4.70	0.79	3.33	0.89	3.27
v_1^*	33.60	33.18	34.10	33.56	28.76	26.99	29.31	27.22
v_2^*	29.37	29.19	29.80	29.49	24.74	23.32	25.20	23.47
v_3^*	34.11	36.07	34.76	36.58	32.49	32.20	33.24	32.62
v_4^*	34.12	36.16	34.76	36.67	32.50	32.30	33.25	32.72
	$n_1 = 30$, $n_2 = 15$							
v_1	118.21	95.99	128.91	103.46	99.36	80.75	106.59	86.75
v_2	-2.52	-2.38	-1.96	-2.44	-2.57	-0.59	-2.79	-0.91
v_3	0.34	-6.51	1.87	-5.77	1.16	-4.79	2.01	-4.50
v_4	0.61	1.71	1.15	1.80	1.36	4.04	1.33	3.93
v_1^*	37.10	34.46	36.69	34.96	32.57	28.90	33.08	29.56
v_2^*	33.79	31.44	33.32	31.86	29.51	26.19	29.96	26.76
v_3^*	37.35	36.44	36.69	37.01	35.11	32.26	35.76	33.11
v_4^*	37.35	36.52	36.69	37.09	35.11	32.34	35.77	33.19
	$n_1 = 40$, $n_2 = 20$							
v_1	110.63	93.73	122.80	101.33	83.90	66.49	88.70	69.98
v_2	-0.31	1.51	1.01	1.61	-0.50	-0.29	-2.14	-0.62
v_3	2.07	-3.82	4.58	-2.97	2.44	-5.32	1.78	-5.20
v_4	2.32	5.14	3.95	5.40	2.61	3.30	1.24	3.04
v_1^*	40.10	37.67	41.93	38.76	31.35	27.05	32.14	27.81
v_2^*	37.28	35.06	39.00	36.08	28.91	24.86	29.62	25.60
v_3^*	40.15	39.47	42.30	40.66	33.11	29.29	34.27	30.24
v_4^*	40.15	39.54	42.30	40.72	33.12	29.35	34.28	30.30

Table 3.2
Relative Efficiencies

$$\left(v_2=\hat{V}^{yg}_{DE}, v_2^*=\hat{V}^{yg}_{HA}, v_3=\hat{\hat{V}}^{ht}_{DE}, v_4=\hat{\hat{V}}^{yg}_{DE}, v_3^*=\hat{\hat{V}}^{ht}_{HA}, v_4^*=\hat{\hat{V}}^{yg}_{HA}\right)$$

Estimators	N=100				N=200			
	$\alpha=0.0$		$\alpha=5.0$		$\alpha=0.0$		$\alpha=5.0$	
	g		g		g		g	
	0	0.5	0	0.5	0	0.5	0	0.5
	$n_2=10\ \ n_1=20$							
v_2	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
v_3	231.54	197.67	224.54	201.50	206.68	87.27	202.62	178.15
v_4	230.68	174.10	229.56	175.01	206.31	167.25	205.14	167.18
v_2^*	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
v_3^*	167.13	129.18	165.46	128.26	163.31	134.98	161.23	133.77
v_4^*	167.09	128.83	165.42	127.91	163.28	134.68	161.21	133.47
	$n_2=15\ \ n_1=30$							
v_2	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
v_3	252.46	191.95	242.16	198.29	213.88	178.36	211.01	184.82
v_4	251.89	189.31	247.28	190.77	213.52	174.92	212.87	174.94
v_2^*	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
v_3^*	153.42	124.89	154.33	123.98	155.52	139.35	153.90	137.42
v_4^*	153.39	124.54	154.30	123.64	155.50	139.04	153.88	137.12
	$n_2=20\ \ n_1=40$							
v_2	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
v_3	269.40	214.62	249.84	221.96	217.21	177.69	216.26	183.20
v_4	268.15	185.82	258.19	186.62	216.65	179.07	218.53	180.86
v_2^*	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
v_3^*	142.19	117.13	137.56	115.85	153.66	142.27	149.37	140.11
v_4^*	142.16	116.83	137.53	115.57	153.64	141.97	149.35	139.83

Table 3.3
Number of Negative Values

$$\left(v_1 = \hat{V}_{DE}^{ht}, v_3 = \hat{\hat{V}}_{DE}^{ht}\right)$$

Estimators	N=100				N=200			
	$\alpha = 0.0$		$\alpha = 5.0$		$\alpha = 0.0$		$\alpha = 5.0$	
	g		g		g		g	
	0.0	0.5	0.0	0.5	0.0	0.5	0.0	0.5
	$n_1 = 20$, $n_2 = 10$							
v_1	32674	31015	33252	31640	31662	29339	32254	29989
v_3	239	2462	1052	2231	287	2960	1058	2638
	$n_1 = 30$, $n_2 = 15$							
v_1	27492	26273	28358	26963	26571	24454	27224	25171
v_3	165	1974	878	1686	209	2335	982	2068
	$n_1 = 40$, $n_2 = 20$							
v_1	30055	27816	30796	28646	26604	23564	27569	24656
v_3	162	1736	1064	1508	220	2138	1057	1728

ACKNOWLEDGEMENT

The authors wish to thank the referees for their valuable suggestions which lead to substantial improvement of the earlier version of the paper.

REFERENCES

1. Asok. C. and Sukhatme, B.V. (1976). On Sampford's procedure of unequal probability sampling without replacement. *J. Amer. Statist. Assoc.*, 71, 912-918
2. Hajek, J. (1971). Discussions of an essay on the logical foundations of survey sampling, part on by D. Basu. In V.P. Godambe & D.A. Sportt (Eds.). *Foundations of Statistical Inference*, Page 326, Toronto, Canada: Holt, Rinchart, Winson.
3. Haziza, D., Hidiroglou, M.A. and Rao, J.N.K. (2011). Comparison of variance estimators in two-phase sampling: An empirical investigation. *Pak. J. Statist.*, 27(4), 477-492.
4. Hidiroglou, M.A., Rao, J.N.K. and Haziza, D. (2009). Variance estimation in two-phase sampling. *Aust. and N.Z.J. Statist. Assoc.*, 47, 663-685.
5. Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, 47, 663-685.
6. Lahiri, D.B. (1951). A method of sample selection providing unbiased ratio estimators. *Bull. Int. Statist. Assoc.*, 33, 133-140.

7. Midzuno, H. (1952). On sampling system with probabilities proportionate to sum of sizes. *Ann. Inst. Statist. Math.*, 3, 99-107.
8. Neyman, J. (1938). Contribution to the theory of sampling human populations. *J. Amer. Statist. Assoc.*, 33, 101-116.
9. Sampford, M.R. (1967). On sampling without replacement with unequal probability selection. *Biometrika*, 67, 639-650.
10. Särndal, C.E., Swensson, B. and Wretman, J. (1992). Model Assisted Survey Sampling. *J. Amer. Statist. Assoc.*, 89, 30-43.
11. Sen, A.R. (1953). On the estimates of variance in sampling with varying probabilities. *J. Ind. Soc. Agri. Statist.*, 5, 119-127.
12. Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *J. Roy. Statist. Soc.*, 15, 235-261.