Journal of the Korean Statistical Society (1998), **27**: *4, pp 459–469*

Unbiased Balanced Half-Sample Variance Estimation in Stratified Two-stage Sampling †

Kyu Seong Kim[1]

ABSTRACT

Balanced half sample method is a simple variance estimation method for complex sampling designs. Since it is simple and flexible, it has been widely used in large scale sample surveys. However, the usual BHS method overestimate the true variance in without replacement sampling and two-stage cluster sampling. Focusing on this point, we proposed an unbiased BHS variance estimator in a stratified two-stage cluster sampling and then described an implementation method of the proposed estimator. Finally, partially BHS design is explained as a tool of reducing the number of replications of the proposed estimator.

Keywords: Variance estimation; Balanced half sample; Partially balanced half sample; Stratified two-stage cluster sampling

1. INTRODUCTION

Complex sampling design has been used to increase the efficiency of the survey in large-scale sample surveys, which contains complex selection and estimation method. Since it can represent well the state of the finite population, complex sample estimate will be more efficient than simple random sample estimate. On the contrary, it gives the difficulty of finding variance estimator of survey estimate. So it has been needed a variance estimation method on complex sampling design, which is simple and flexible. As results of many researcher's efforts some methods were proposed and the balanced half sample(BHS) method is a widely used one among them.

BHS method was originated by McCarthy(1966, 1969) and his work was followed by a number of authors. Kish and Frankel(1970, 1974) suggested BHS

†This paper was supported by the University of Seoul Research Fund, 1997

[1]Department of Computer Science and Statistics, The University of Seoul, Seoul, 130-743,Korea

variance estimator of nonlinear parameter in stratified sampling and compared the efficiency by a simulation. Krewski(1978) studied the stability of BHS variance estimator and Krewski and Rao(1981) made asymptotic inference of variance estimators over some resamples. Lee(1973) used partially BHSs for variance estimation to overcome the difficulty of increasing the number of BHSs when the number of strata is large. A nice summary for several variance estimation methods based on resamples was accomplished by Wolter(1985). After then, many results to complement the weakness of BHS mathod are presented. For example, Dippo and Wolter(1984) studied grouped balanced half sample(GBHS) method, which is useful when more than two PSUs are selected in a stratum, and Rao and Shao(1996) suggested grouped balanced half sample(RGBHS) method. In their paper, Rao and Shao showed that their method is asymptotically correct as well as Dippo and Wolter's method is asymptotically incorrect.

It is a merit that BHS method provides a simple variance estimation method for complex sampling designs. However, there is a demerit that BHS method usually overestimate the true variance in without replacement sampling and multistage cluster sampling. Focusing on this point, we have studied for finding unbiased BHS variance estimator over stratified two-stage cluster sampling in this paper. In section 2, we review a basic BHS method in stratified two-stage cluster sampling and show that the usual BHS variance estimator is unbiased when with replacement sampling is used at the first stage. Section 3 proposes an unbiased BHS variance estimator when without replacemant sampling, especially inclusion probability proportional to size sampling(πPS), is used and describes an implementation method. In addition, we introduce partially BHS design and illustrate the method of reducing the number of replications. Finally concluding remarks are discussed in section 4.

2. BALANCED HALF SAMPLE METHOD

The population is divided into L strata with stratum h containing N_h primary sampling units(PSUs) and the (hi)th PSU contains M_{hi} secondary sampling units(SSUs). Let $N = \sum_{h=1}^{L} N_h$ be the total number of PSUs and $M = \sum_{h=1}^{L} \sum_{i=1}^{N_h} M_{hi}$ the total number of SSUs. By y_{hij} we denote the value of (hij)th SSU, which is assumed to be observed without measurement errors. For estimating the population total, $Y = \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{j=1}^{M_{hi}} y_{hij}$, select two PSUs from each stratum, which will be denoted by $(h1)$, $(h2)$ and choose m_{hi} SSUs from (hi)th PSU. We will denote a subsample of SSUs by s_{hi}. Then a general estimator of Y

may be written as

$$\hat{Y} = g(\bar{y}_{hi} \ : \ h = 1, ..., L, i = 1, 2) \tag{2.1}$$

where $\bar{y}_{hi} = \sum_{j \in s_{hi}} y_{hij}/m_{hi}$ and g is a smooth function usually.

In BHS method, a half sample is defined by selecting randomly only one PSU from the original sample for each stratum, then 2^L half samples are possible. Similar to a general estimator $\hat{Y}$, define a half sample estimator based on αth half sample as

$$\hat{Y}_\alpha = g((1 + \delta_h^{(\alpha)})\bar{y}_{h1}, \ (1 - \delta_{h2}^{(\alpha)})\bar{y}_{h2} \ : \ h = 1, ..., L) \tag{2.2}$$

where

$$\delta_h^{(\alpha)} = \begin{cases} 1 & \text{if } (h1) \in \alpha\text{th half sample} \\ -1 & \text{if } (h2) \in \alpha\text{th half sample} \end{cases}$$

A BHS variance estimator based on k BHSs is then defined by

$$v_k(\hat{Y}) = \frac{1}{k}\sum_{\alpha=1}^{k}(\hat{Y}_\alpha - \hat{Y})^2, \tag{2.3}$$

where k BHSs are satisfied the following conditions,

$$\sum_{\alpha=1}^{k} \delta_h^{(\alpha)}\delta_{h'}^{(\alpha)} = 0 \ \text{ for all } \ h \neq h'. \tag{2.4}$$

To investigate the property of the BHS variance estimator, $v_k(\hat{Y})$, we first consider probability proportional to size sampling with replacement(PPSWR) at the first stage and take any sampling design at the second stage. Let p_{hi} is the probability of selecting (hi)th PSU, then the usual unbiased estimator for the population total is given by $\hat{Y} = \sum_{h=1}^{L}(\hat{Y}_{h1}/2p_{h1} + \hat{Y}_{h2}/2p_{h2})$, where $\hat{Y}_{hi}$ $(i = 1, 2)$ is an unbiased estimator for the population total of the (hi)th PSU. With σ_{hi}^2 the variance of $\hat{Y}_{hi}$ at the second stage, the variance of $\hat{Y}$ is given by

$$Var\{\hat{Y}\} = \frac{1}{4}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{j(\neq i)}^{N_h} p_{hi}p_{hj}(\frac{Y_{hi}}{p_{hi}} - \frac{Y_{hj}}{p_{hj}})^2 + \sum_{h=1}^{L}\sum_{i=1}^{N_h}\frac{\sigma_{hi}^2}{2p_{hi}}. \tag{2.5}$$

The following theorem says that the usual BHS method provides an unbiased estimator in stratified two-stage cluster sampling, in which PPSWR is used at the first stage.

Theorem 2.1. *In a stratified two-stage cluster sampling, in which PSU is selected by PPSWR and SSU by any sampling design, the following BHS variance estimator is unbiased for the variance of* $\hat{Y}$ *:*

$$v_k(\hat{Y}) = \frac{1}{k}\sum_{\alpha=1}^{k}(\hat{Y}_\alpha - \hat{Y})^2,$$

where

$$\hat{Y}_\alpha = \sum_{h=1}^{L}\left(\frac{(1+\delta_h^{(\alpha)})}{2}\frac{\hat{Y}_{h1}}{p_{h1}} + \frac{(1-\delta_h^{(\alpha)})}{2}\frac{\hat{Y}_{h2}}{p_{h2}}\right) \tag{2.6}$$

is the α*th half sample estimator.*

Proof: From $\hat{Y}_\alpha = \hat{Y} + \sum_{h=1}^{L}(\hat{Y}_{h1}/p_{h1} - \hat{Y}_{h2}/p_{h2})\delta_h^{(\alpha)}/2$ and $\sum_{\alpha=1}^{k}\delta_h^{(\alpha)}\delta_{h'}^{(\alpha)} = 0$ for two different strata h and h', we get

$$\begin{aligned} v_k(\hat{Y}) &= \frac{1}{k}\sum_{\alpha=1}^{k}\{\sum_{h=1}^{L}(\frac{\hat{Y}_{h1}}{p_{h1}} - \frac{\hat{Y}_{h2}}{p_{h2}})\frac{\delta_h^{(\alpha)}}{2}\}^2 \\ &= \frac{1}{4}\sum_{h=1}^{L}(\frac{\hat{Y}_{h1}}{p_{h1}} - \frac{\hat{Y}_{h2}}{p_{h2}})^2, \end{aligned}$$

which is an usual unbiased variance estimator of $\hat{Y}$, so that $v_k(\hat{Y})$ is unbiased. □

Corollary 2.1. *If a subsample* s_{hi} *is selected by simple ramdom sampling without repalcement(SRSWOR) in the above sampling procedure, then* α*th half sample estimator is given by*

$$\hat{Y}_\alpha = \sum_{h=1}^{L}\left(\frac{(1+\delta_h^{(\alpha)})}{2}\frac{\hat{Y}_{h1}}{p_{h1}} + \frac{(1-\delta_h^{(\alpha)})}{2}\frac{\hat{Y}_{h2}}{p_{h2}}\right) \text{ with } \hat{Y}_{hi} = \frac{M_{hi}}{m_{hi}}\sum_{j\in s_{hi}} y_{hij} \tag{2.7}$$

and the BHS variance estimator $v_k(\hat{Y})$ *is unbiased for the variance of* $\hat{Y}$.

Now we extend the BHS method to a stratified two-stage cluster sampling, in which PSUs are selected by without replacement sampling, especially by inclusion probability proportional to size sampling (πPS). Since BHS variance estimator is unbiased when PSUs are selected by PPSWR at the first stage as shown Theorem 2.1, the usual BHS variance estimator cannot help overestimating the variance under a without replacement sampling procedure at the first stage.

Specifically, we suppose that two PSUs are selected by πPS from each stratum and SSUs by any sampling design from the selected PSU. Then we can get an usual estimator of population total as $\hat{Y} = \sum_{h=1}^{L}(\hat{Y}_{h1}/\pi_{h1} + \hat{Y}_{h2}/\pi_{h2})$, where π_{hi} is the inclusion probability of (hi)th PSU and is equal to $2p_{hi}$ in πPS sampling. The variance of $\hat{Y}$ is then represented as

$$Var\{\hat{Y}\} = \sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}(\pi_{hi}\pi_{hj} - \pi_{hij})(\frac{Y_{hi}}{\pi_{hi}} - \frac{Y_{hj}}{\pi_{hj}})^2 + \sum_{h=1}^{L}\sum_{i=1}^{N_h}\frac{\sigma_{hi}^2}{\pi_{hi}}, \tag{2.8}$$

Under this sampling procedure, if we take $v_k(\hat{Y})$ based on $\hat{Y}_\alpha = \sum_{h=1}^{L}((1+\delta_h^{(\alpha)})$ $\hat{Y}_{h1}/\pi_{h1}$ $+(1-\delta_h^{(\alpha)})$ $\hat{Y}_{h2}/\pi_{h2})$ as a variance estimator, then serious bias problem may happen. The quantity of bias is obtained by

$$B_1 = \sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}(2\pi_{hij} - \pi_{hi}\pi_{hj})(\frac{Y_{hi}}{\pi_{hi}} - \frac{Y_{hj}}{\pi_{hj}})^2. \tag{2.9}$$

Since B_1 contains a variation of PSUs, it may be large in heterogeneous clustered population. So the bias should be removed by adjusting the BHS variance estimator.

To make an unbiased BHS variance estimator, at first, the following simply modified half sample estimator such as

$$\hat{Y}_\alpha^* = \hat{Y} + \sum_{h=1}^{L} W_h^*(\frac{\hat{Y}_{h1}}{\pi_{h1}} - \frac{\hat{Y}_{h2}}{\pi_{h2}})\delta_h^{(\alpha)}, \tag{2.10}$$

where $W_h^* = ((\pi_{h1}\pi_{h2} - \pi_{h12})/\pi_{h12})^{1/2}$ can be considered. Then BHS variance estimator is reduced to

$$v_k^*(\hat{Y}) = \sum_{h=1}^{L}\frac{\pi_{h1}\pi_{h2} - \pi_{h12}}{\pi_{h12}}(\frac{\hat{Y}_{h1}}{\pi_{h1}} - \frac{\hat{Y}_{h2}}{\pi_{h2}})^2. \tag{2.11}$$

Unfortunately this estimator does not estimate the second stage variance unbiasedly, it is biased estimator for the varaince in (2.8). Here the bias of $v_k^*(\hat{Y})$ is given by

$$B_2 = \sum_{h=1}^{L}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}(\pi_{hi}\pi_{hj} - \pi_{hij})(\frac{\sigma_{hi}^2}{\pi_{hi}^2} + \frac{\sigma_{hj}^2}{\pi_{hj}^2}) - \sum_{h=1}^{L}\sum_{i=1}^{N_h}\frac{\sigma_{hi}^2}{\pi_{hi}}. \tag{2.12}$$

Therefore more precise adjustment factors should be considered in BHS method for estimating the second stage variance as well as the total variance unbiasedly.

In the next section, we will propose an unbiased BHS variance estimator under our sampling procedure.

3. UNBIASED BALANCED HALF SAMPLE METHOD IN STRATIFIED TWO-STAGE SAMPLING

3.1. Unbiased BHS method

For constructing unbiased BHS variance estimator, it is necessary to estimate unbiasedly the second stage variance as well as the first stage variance in the context of BHS method. One method is to add some additional terms to $v_k^*(\hat{Y})$ in order to estimate the second stage variance unbiasedly.

In the αth replication, we introduce two dummy variables $\eta_h^{(\alpha)}$ and $\kappa_h^{(\alpha)}$ having one of two values, $+1$ or -1, such that $\delta_h^{(\alpha)}$, $\eta_h^{(\alpha)}$ and $\kappa_h^{(\alpha)}$ are mutually orthogonal, i.e.

$$\sum_{\alpha=1}^{k} \delta_h^{(\alpha)} \delta_{h'}^{(\alpha)} = \sum_{\alpha=1}^{k} \eta_h^{(\alpha)} \eta_{h'}^{(\alpha)} = \sum_{\alpha=1}^{k} \kappa_h^{(\alpha)} \kappa_{h'}^{(\alpha)} = 0, \quad h \neq h' \tag{3.1}$$

and

$$\sum_{\alpha=1}^{k} \delta_h^{(\alpha)} \eta_{h'}^{(\alpha)} = \sum_{\alpha=1}^{k} \delta_h^{(\alpha)} \kappa_{h'}^{(\alpha)} = \sum_{\alpha=1}^{k} \eta_h^{(\alpha)} \kappa_{h'}^{(\alpha)} = 0, \quad h = h' \text{ or } h \neq h'.$$

Now we define the αth half sample estimator as

$$\hat{Y}_\alpha^{**} = \hat{Y} + \sum_{h=1}^{L} (W_h^* (\frac{\hat{Y}_{h1}}{\pi_{h1}} - \frac{\hat{Y}_{h2}}{\pi_{h2}}) \delta_h^{(\alpha)} + \frac{\hat{\sigma}_{h1}}{\sqrt{\pi_{h1}}} \eta_h^{(\alpha)} + \frac{\hat{\sigma}_{h2}}{\sqrt{\pi_{h2}}} \kappa_h^{(\alpha)}).$$

Then BHS variance estimator based on k balanced half sample estimators can be represented as

$$\begin{aligned} v_k^{**}(\hat{Y}) &= \frac{1}{k} \sum_{\alpha=1}^{k} (\hat{Y}_\alpha^{**} - \hat{Y})^2 \\ &= \sum_{h=1}^{L} \frac{\pi_{h1}\pi_{h2} - \pi_{h12}}{\pi_{h12}} (\frac{\hat{Y}_{h1}}{\pi_{h1}} - \frac{\hat{Y}_{h2}}{\pi_{h2}})^2 + \sum_{h=1}^{L} (\frac{\hat{\sigma}_{h1}^2}{\pi_{h1}} + \frac{\hat{\sigma}_{h2}^2}{\pi_{h2}}), \end{aligned}$$

which is the usual unbiased estimator of the variance in (2.8). This means that the variacne in (2.8) can be unbiasedly estimated by means of BHS method under stratified two-stage sampling, in which PSUs are selected by πPS sampling. This result is formally stated in the following theorem.

Theorem 3.1. *Suppose a stratified two-stage cluster sampling, in which two PSUs are selected by πPS and SSUs by any sampling design. Based on the k BHSs, we define the αth half sample estimator as*

$$\hat{Y}_\alpha^{**} = \hat{Y} + \sum_{h=1}^{L}(W_h^*(\frac{\hat{Y}_{h1}}{\pi_{h1}} - \frac{\hat{Y}_{h2}}{\pi_{h2}})\delta_h^{(\alpha)} + \frac{\hat{\sigma}_{h1}}{\sqrt{\pi_{h1}}}\eta_h^{(\alpha)} + \frac{\hat{\sigma}_{h2}}{\sqrt{\pi_{h2}}}\kappa_h^{(\alpha)}),\ (\alpha = 1, ..., k), \quad (3.2)$$

where $\delta_h^{(\alpha)}$, $\eta_h^{(\alpha)}$ and $\kappa_h^{(\alpha)}$ are mutually orthogonal. Then the following BHS variance estimator,

$$v_k^{**}(\hat{Y}) = \frac{1}{k}\sum_{\alpha=1}^{k}(\hat{Y}_\alpha^{**} - \hat{Y})^2, \quad (3.3)$$

is unbiased for the variance of $\hat{Y}$.

Note 3.1. In each stratum, a πPS sample of size 2 can be obtained by Sampford's sampling scheme(1967) as follows :

Step 1) Select the first PSU with probability p_{hi}.

Step 2) Select the second PSU with probability proportional to $p_{hi}/(1-2p_{hi})$.

Step 3) If two PSU are distinct, accept them as a sample. If not, reject two PSU and the process is restarted until two PSUs are distinct.

Note 3.2. In the Sampford's sampling procedure, the second order inclusion probability is given by

$$\pi_{hij} = \frac{2p_{hi}p_{hj}}{D_h}\frac{1-p_{hi}-p_{hj}}{(1-2p_{hi})(1-2p_{hj})}, \quad \text{where } D_h = \frac{1}{2}(1 + \sum_{i=1}^{N_h}\frac{p_{hi}}{1-2p_{hi}}).$$

In addition, $\pi_{hij} \leq \pi_{hi}\pi_{hj}$ holds for all (hij).

In BHS method, the number of replication k for given strata number L should be determined to satisfy BHS condition, say $\sum_{\alpha=1}^{k}\delta_h^{(\alpha)}\delta_{h'}^{(\alpha)} = 0,\ h \neq h'$. Plackett and Burman (1946) have suggested methods for constructing $k \times k$ orthogonal matrices, where k is a multiple of 4. In their matrices, minimal set of BHSs are given such that $L < k \leq L + 4$. For example, if $L = 6$, then minimal set of replication is given by 8.

An unbiased condition of the proposed estimator $v_k^{**}(\hat{Y})$ is that $\delta_h^{(\alpha)}, \eta_h^{(\alpha)}$ and $\kappa_h^{(\alpha)}$ are mutually orthogonal. To satisfy this condition, three times of replications, say $3k$, are needed. For example, when there are three strata, minimal

Table 3.1: A BHS design for the proposed BHS variance estimator when $L = 3$ strata

Replicate(α)	$\delta_1^{(\alpha)}$	$\delta_2^{(\alpha)}$	$\delta_3^{(\alpha)}$	$\eta_1^{(\alpha)}$	$\eta_2^{(\alpha)}$	$\eta_3^{(\alpha)}$	$\kappa_1^{(\alpha)}$	$\kappa_2^{(\alpha)}$	$\kappa_3^{(\alpha)}$
1	+1	+1	+1	+1	+1	+1	+1	+1	+1
2	−1	+1	−1	+1	+1	+1	−1	−1	−1
3	−1	−1	+1	−1	+1	+1	+1	−1	−1
4	+1	−1	−1	+1	−1	+1	+1	+1	−1
5	−1	+1	−1	−1	+1	−1	+1	+1	+1
6	−1	−1	+1	−1	−1	+1	−1	+1	+1
7	−1	−1	−1	+1	−1	−1	+1	−1	+1
8	+1	−1	−1	−1	+1	−1	−1	+1	−1
9	+1	+1	−1	−1	−1	+1	−1	−1	+1
10	+1	+1	+1	−1	−1	−1	+1	−1	−1
11	−1	+1	+1	+1	−1	−1	−1	+1	−1
12	+1	−1	+1	+1	+1	−1	−1	−1	+1

number of replication is 4 for the usual BHS method and 12 for the proposed BHS method. A BHS design for the proposed BHS variance estimator when $L = 3$ is given in below Table 3.1.

3.2 Partially balanced half sample (PBHS) method

Although the proposed BHS method gives an unbiased variance estimator, it is unsatisfactory for sampling designers to need much more replications than the usual BHS method. It is because the cost of processing may be high when the number of strata is large.

To reduce the total number of replication, we introduce a partially balanced half sample (PBHS) design. Let $\mathbf{D}$ be a $k \times L$ BHS design, then 3-order PBHS design, denoted by $\mathbf{PD}$, is defined as

$$\mathbf{PD} = (\mathbf{D}, \mathbf{D}, \mathbf{D}), \tag{3.4}$$

which has $k \times 3L$ dimensions so that the number of replication is the same that of the usual BHS design $\mathbf{D}$. For example, 3-order PBHS design where $L = 3$ is presented in the below Table 3.2.

In PBHS design, mutually orthogonal condition does not hold any more since

Table 3.2: 3-order PBHS design for $L = 3$ strata

Replicate(α)	$\delta_1^{(\alpha)}$	$\delta_2^{(\alpha)}$	$\delta_3^{(\alpha)}$	$\eta_1^{(\alpha)}$	$\eta_2^{(\alpha)}$	$\eta_3^{(\alpha)}$	$\kappa_1^{(\alpha)}$	$\kappa_2^{(\alpha)}$	$\kappa_3^{(\alpha)}$
1	+1	+1	+1	+1	+1	+1	+1	+1	+1
2	−1	+1	−1	−1	+1	−1	−1	+1	−1
3	−1	−1	+1	−1	−1	+1	−1	−1	+1
4	+1	−1	−1	+1	−1	−1	+1	−1	−1

the following equations does hold at the corresponding stratum,

$$\sum_{\alpha=1}^{k} \delta_h^{(\alpha)} \eta_{h'}^{(\alpha)} = \sum_{\alpha=1}^{k} \delta_h^{(\alpha)} \kappa_{h'}^{(\alpha)} = \sum_{\alpha=1}^{k} \eta_h^{(\alpha)} \kappa_{h'}^{(\alpha)} = k, \quad h = h'. \tag{3.5}$$

This means that PBHS method can reduce the total number of replications to the 1/3 times as well as it fails the condition of orthogonality of $\delta_h^{(\alpha)}, \eta_h^{(\alpha)}$ and $\kappa_h^{(\alpha)}$. Based on PBHS design, the bias of the estimator $v_k^{**}(\hat{Y})$ is given by

$$B_3 = \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} \pi_{hij} (W_{hij}^* (\frac{Y_{hi}}{\pi_{hi}} - \frac{Y_{hj}}{\pi_{hj}})(\frac{\sigma_{hi}}{\sqrt{\pi_{hi}}} + \frac{\sigma_{hj}}{\sqrt{\pi_{hj}}}) + \frac{\sigma_{hi}\sigma_{hj}}{\sqrt{\pi_{hi}\pi_{hj}}}) \tag{3.6}$$

where $W_{hij}^* = ((\pi_{hi}\pi_{hj} - \pi_{hij})/\pi_{hij})^{1/2}$.

For the comparison of B_2 and B_3, we make two simple but reasonable assumptions such that for each h (i) $\sigma_{hi} = \sigma_{hj} = \sigma_h$ and (ii) $(\pi_{hi}\pi_{hj} - \pi_{hij})/\pi_{hij} \le A_h$ for some positive A_h. Then upper bounds of B_2 and B_3 can be obtained by

$$B_2 \le \sum_{h=1}^{L} \sigma_h^2 (\tau_h (2A_h - 1) - A_h N_h), \quad \text{where} \quad \tau_h = \sum_{i=1}^{N_h} \frac{1}{\pi_{hi}} \tag{3.7}$$

and

$$B_3 \le \sum_{h=1}^{L} \sigma_h^2 (\frac{\phi_h^2}{2} - 1), \quad \text{where} \quad \phi_h = \sum_{i=1}^{N_h} \sqrt{\pi_{hi}}. \tag{3.8}$$

To clarify the meaning of B_2 and B_3, we make two examples as follows.

Example 3.1: In two-stage cluster sampling, let $p_{hi} = M_{hi}/M_h$, $M_h = \sum_{i=1}^{N_h} M_{hi}$ and assume that L is fixed and $\max(M_{hi})$ is bounded. Then in each stratum, $\tau_h = O(N_h^2)$, $A_h = O(1)$, and $\phi_h^2 = O(N_h)$ as $N_h \to \infty$. Hence we get

$$B_2 \le O(N_h^2) \quad \text{and} \quad B_3 \le O(N_h).$$

This result says that $v_k^{**}(\hat{Y})$ based on PBHS is more efficient than $v_k^*(\hat{Y})$ in view of bias when the stratum size is large.

Example 3.2: When PSUs are selected by SRSWOR, we have

$$B_2 = \frac{1}{2}\sum_{h=1}^{L}\sigma_h^2(N_h^2 - 6N_h + 4), \quad \text{and} \quad B_3 = \sum_{i=1}^{N_h}\sigma_h^2(N_h - 1).$$

Furthermore, it is obvious that $B_2 > B_3$ when $N_h \geq 8$ for all stratum h.

4. CONCLUSION

The study of variance estimation on the complex sampling design has been receiving considerable attention, because it is very important but difficult. The BHS method, which does not depend on the specified sampling design, has been used widely because of its simplicity and flexibility. Contrary to this merit, there is a demerit that BHS method result in overestimating the variance in without replacement sampling and multi-stage sampling. In this paper, we proposed an unbiased BHS method in stratified two-stage cluster sampling, in which PSU is selected by πPS and described an implementatin method of the proposed estimator. Finally, we illustrated partially BHS design to reduce the number of replications.

REFERENCES

Dippo, C.S. and Wolter, K.M. (1984). "A comparison of variance estimators using the Taylor series approximation," *Proceeding of the survey research section*, American Statistical Association.

Kish, L. and Frankel, M.R. (1970). "Balanced repeated replication for standard errors," *Journal of the American Statistical Association*, **65**, 1071-1094.

Kish, L. and Frankel, M.R. (1974). "Inference from complex samples," *Journal of Royal Statistical Society*, **B36**, 1-37.

Krewski, D. (1978). "On the stability of some replication variance estimators in the linear case," *Journal of Statistical Planning and Inference*, **2**, 45-51.

Krewski, D. and Rao, J.N.K. (1981). "Inference from stratified samples : properties of linearization, Jackknife and balanced repeated replication methods," *Annals of Statistics*, **9**, 1010-1019.

Lee, K.H. (1973). "Variance estimation in stratified sampling," *Journal of the American Statistical Association*, **68**, 336-342.

McCarthy, P.J. (1966). "Replication: An approach to the analysis of data from complex surveys," *Vital and Health Statistics*, Series 2, No. 14, National Center for Health Statistics, Public Health Service, Washington. D.C.

McCarthy, P.J. (1969). "Pseudo-replication : Half samples," *International Statistical Review*, **37**, 239-264.

Plackett, R.L. and Burman, J.P. (1946). "The design of optimum multifactorial experiments," *Biometrika*, **33**, 305-325.

Rao, J.N.K. and Shao, J. (1996). "On balanced half-sample variance estimation in stratified random sampling," *Journal of the American Statistical Association*, **91**, 343-348.

Sampford, M.R. (1967). "On sampling without replacement with unequal probabilities of selection," *Biometrika*, **54**, 499-513.

Wolter, K.M. (1985). *Introduction to Variance Estimation*. Springer-Verlag, New York.