An Alternative Estimator in Finite Population Sampling on Successive Occasions

Bongkoj Wibultananun [1], Prachoom Suwattee [2]

[1] National Statistical Office, Bangkok (wbongkoj@yahoo.co.th)
[2] School of Applied Statistics, National Institute of Development Administration, Bangkok (prachoom@as.nida.ac.th)

Abstract

This study aims to propose an alternative estimator of the population mean in sampling on successive occasions from a fixed finite population, and to compare this estimator with two other estimators which are an unbiased estimator and a conventional estimator.

The alternative estimator is always consistent, whereas the others are consistent for some weights used in the estimation. The conventional estimator and the alternative estimator are biased if the regression coefficient used in the estimation is not a pre-assigned constant, where the absolute value of the bias of the alternative estimator is less than that of the other. The variance of the unbiased estimator is more than the variances of the others. For an interval of weights used in the estimation, the variance of the alternative estimator is not less than that of the conventional estimator. Using improper weights, the variance of the unbiased estimator and the mean square error of the conventional estimator may be very high, whereas the mean square error of the alternative estimator is slightly different from the minimum.

1. Introduction

Sampling on successive occasions is the organized process of eliminating some units from the sample and adding new units to the sample as time advances. It may be used in the surveys which the same population is sampled and the same study variable is measured repeatedly on different occasions.

In this study, we assume that the units in the population remain unchanged in time, but the values of the study variable may be changed. Let N be the population size on every occasion. For $k = 1, 2, 3, \ldots, t$; let s_k be the sample on the k^{th} occasion, of size n_k. On the first occasion, n_1 units are drawn from the population by simple random sampling without replacement. On the k^{th} occasion, $k = 2, 3, 4, \ldots, t$; n_{km} units are drawn from the n_{k-1} units of s_{k-1} by simple random sampling without replacement, and n_{ku} units are drawn from the remaining $N - n_{k-1}$ units of the population also by simple random sampling without replacement. So $n_k = n_{km} + n_{ku}$ for $k = 2, 3, 4, \ldots, t$.

Since there exists a relationship between data attached to a unit of the population on two or more successive occasions, it is possible to use the data contained in the previous samples to improve the estimate of a population parameter on the current occasion. The first attempt was made by Jessen [2] for two successive occasions. He combined two estimates for estimating the population mean on the second occasion; one is the mean based on the *unmatched* portion of the sample on the second occasion, and the other is a double-sampling regression estimate from the *matched* portion. The initial results of Jessen [2] have been extended to more than two successive occasions by Yates [4], Patterson [3], Cochran [1], and several others. However, the theory has been almost confined to large population.

In this study, three estimators of the population mean on the current occasion, which can be applied in finite population sampling on successive occasions, are presented. The consistency, biases and variances, as well as mean square errors of these estimators are compared.

2. Estimators of the Population Mean

Let t be the current occasion, $t \geq 2$. For $k = 1, 2, 3, \ldots, t$, the y value on the k^{th} occasion of unit i is denoted by y_{ki}, and we let $\overline{Y}_k$ be the population mean on the k^{th} occasion, that is,

$$\overline{Y}_k = \frac{\sum_{i=1}^{N} y_{ki}}{N}, \tag{1}$$

and $\overline{y}_k$ be the mean on the k^{th} occasion based on the whole sample, that is,

$$\overline{y}_k = \frac{\sum_{i=1}^{n_k} y_{ki}}{n_k}. \tag{2}$$

For $k = 2,3,4,\ldots,t$, let $\overline{y}_{km}$ be the mean on the k^{th} occasion based on the matched portion of the sample on the k^{th} occasion with the sample on the $(k-1)^{\text{th}}$ occasion, that is,

$$\overline{y}_{km} = \frac{\sum_{i=1}^{n_{km}} y_{ki}}{n_{km}}, \tag{3}$$

$\overline{y}_{ku}$ be the mean on the k^{th} occasion based on the unmatched portion of the sample on the k^{th} occasion, that is,

$$\overline{y}_{ku} = \frac{\sum_{i=1}^{n_{ku}} y_{ki}}{n_{ku}}, \tag{4}$$

and $\overline{y}_{k-1,m*}$ be the mean on the $(k-1)^{\text{th}}$ occasion based on the matched portion of the sample on the k^{th} occasion with the sample on the $(k-1)^{\text{th}}$ occasion, that is,

$$\overline{y}_{k-1,m*} = \frac{\sum_{i=1}^{n_{km}} y_{k-1,i}}{n_{km}}. \tag{5}$$

On the first occasion, $\overline{y}_1$ is usually used as the estimator of $\overline{Y}_1$. On the t^{th} occasion, $t \ge 2$, $\overline{Y}_t$ can be estimated by

$$\hat{\overline{Y}}_t' = W_t'\overline{y}_{tu} + (1-W_t')\overline{y}_{tm}, \tag{6}$$

or

$$\hat{\overline{Y}}_t'' = W_t''\overline{y}_{tu} + (1-W_t'')\left(\overline{y}_{tm} + b_t(\overline{y}_{t-1} - \overline{y}_{t-1,m*})\right), \tag{7}$$

where W_t' and W_t'' are nonnegative constant weights, $0 \le W_t', W_t'' \le 1$, and b_t is a regression coefficient of y_t on y_{t-1}.

An alternative estimator of $\overline{Y}_t$, $t \ge 2$, that we propose is

$$\hat{\overline{Y}}_t = \frac{1}{n_{t-1}+n_{tu}}\left(n_{tu}\overline{y}_{tu} + n_{tm}\overline{y}_{tm} + (n_{t-1}-n_{tm})\left(W_t\overline{y}_{tu} + (1-W_t)\left(\overline{y}_{tm} + b_t(\overline{y}_{t-1,d} - \overline{y}_{t-1,m*})\right)\right)\right), \tag{8}$$

where W_t is a nonnegative constant weight, $0 \le W_t \le 1$, and $\overline{y}_{t-1,d}$ is the mean on the $(t-1)^{\text{th}}$ occasion based on the $n_{t-1} - n_{tm}$ units in the sample on the $(t-1)^{\text{th}}$ occasion which are discarded on the t^{th} occasion, that is,

$$\overline{y}_{t-1,d} = \frac{n_{t-1}\overline{y}_{t-1} - n_{tm}\overline{y}_{t-1,m*}}{n_{t-1} - n_{tm}}. \tag{9}$$

We may write $\hat{\overline{Y}}_t$ in the form

$$\hat{\overline{Y}}_t = \frac{1}{n_{t-1}+n_{tu}}\left(W_t\left((n_{t-1}+n_{tu}-n_{tm})\overline{y}_{tu} + n_{tm}\overline{y}_{tm}\right) + (1-W_t)\left(n_{tu}\overline{y}_{tu} + n_{t-1}\left(\overline{y}_{tm} + b_t(\overline{y}_{t-1} - \overline{y}_{t-1,m*})\right)\right)\right). \tag{10}$$

We find that, the expected values of $\hat{\bar{Y}}_t'$, $\hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$ are

$$E(\hat{\bar{Y}}_t') = \bar{Y}_t \,, \tag{11}$$

$$E(\hat{\bar{Y}}_t'') = \bar{Y}_t + (1 - W_t'')C(b_t, \bar{y}_{t-1} - \bar{y}_{t-1,m^*}) \,, \tag{12}$$

$$E(\hat{\bar{Y}}_t) = \bar{Y}_t + (1 - W_t)\frac{n_{t-1}}{n_{t-1} + n_{tu}} C(b_t, \bar{y}_{t-1} - \bar{y}_{t-1,m^*}) \,, \tag{13}$$

respectively, where $C(b_t, \bar{y}_{t-1} - \bar{y}_{t-1,m^*})$ is the covariance of b_t and $\bar{y}_{t-1} - \bar{y}_{t-1,m^*}$.

When b_t is the population regression coefficient of y_t on y_{t-1}, we find that the variances of $\hat{\bar{Y}}_t'$, $\hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$ are

$$V(\hat{\bar{Y}}_t') = S_t^2 \left(\frac{W_t'^2}{n_{tu}} + \frac{(1 - W_t')^2}{n_{tm}} - \frac{1}{N} \right), \tag{14}$$

$$V(\hat{\bar{Y}}_t'') = S_t^2 \left(\frac{W_t''^2}{n_{tu}} + (1 - W_t'')^2 \left(\frac{1 - \rho_t^2}{n_{tm}} + \frac{\rho_t^2}{n_{t-1}} \right) - \frac{1}{N} \right), \tag{15}$$

$$\begin{aligned} V(\hat{\bar{Y}}_t) = {} & \frac{S_t^2}{n_{t-1} + n_{tu}} \left(1 - \frac{n_{t-1} + n_{tu}}{N} \right) + \frac{S_t^2 (n_{t-1} - n_{tm})}{n_{tu} n_{tm} (n_{t-1} + n_{tu})^2} \Big((1 - \rho_t^2) n_{tu} n_{tm} + W_t^2 n_{tm} (n_{t-1} - n_{tm} + \rho_t^2 n_{tu}) \\ & + (1 - W_t)^2 (1 - \rho_t^2) n_{tu} (n_{t-1} - n_{tm}) \Big), \end{aligned} \tag{16}$$

respectively, where

$$S_t^2 = \frac{\sum_{i=1}^{N} (y_{ti} - \bar{Y}_t)^2}{N - 1} \,, \tag{17}$$

and

$$\rho_t = \frac{\sum_{i=1}^{N} (y_{ti} - \bar{Y}_t)(y_{t-1,i} - \bar{Y}_{t-1})}{\sqrt{\sum_{i=1}^{N} (y_{ti} - \bar{Y}_t)^2 \sum_{i=1}^{N} (y_{t-1,i} - \bar{Y}_{t-1})^2}} \,. \tag{18}$$

In practice, if b_t is not pre-assigned we may use

$$b_t = \frac{\sum_{i=1}^{n_{tm}} (y_{ti} - \bar{y}_{tm})(y_{t-1,i} - \bar{y}_{t-1,m*})}{\sum_{i=1}^{n_{tm}} (y_{t-1,i} - \bar{y}_{t-1,m*})^2} \,, \tag{19}$$

which is the least squares estimate of the population regression coefficient of y_t on y_{t-1}. The variances of $\hat{\bar{Y}}_t'$, $\hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$ may be estimated by replacing S_t^2 and ρ_t by $\hat{S}_t^2$ and $\hat{\rho}_t$, respectively, where

$$\hat{S}_t^2 = \frac{\sum_{i=1}^{n_t} (y_{ti} - \bar{y}_t)^2}{n_t - 1} \,, \tag{20}$$

and

$$\hat{\rho}_t = \frac{\sum_{i=1}^{n_{tm}} (y_{ti} - \bar{y}_{tm})(y_{t-1,i} - \bar{y}_{t-1,m^*})}{\sqrt{\sum_{i=1}^{n_{tm}} (y_{ti} - \bar{y}_{tm})^2 \sum_{i=1}^{n_{tm}} (y_{t-1,i} - \bar{y}_{t-1,m^*})^2}} . \quad (21)$$

The minimum variances of $\hat{\bar{Y}}_t', \hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$ can be obtained by using the proper W_t', W_t'' and W_t which minimize (14), (15) and (16), respectively. We find that, the proper W_t', W_t'' and W_t are

$$W_t' = \frac{n_{tu}}{n_t} , \quad (22)$$

$$W_t'' = \frac{n_{tu}\left((1-\rho_t^2)n_{t-1} + \rho_t^2 n_{tm}\right)}{n_{t-1}n_{tm} + n_{tu}\left((1-\rho_t^2)n_{t-1} + \rho_t^2 n_{tm}\right)} , \quad (23)$$

$$W_t = \frac{(1-\rho_t^2)n_{tu}(n_{t-1} - n_{tm})}{n_{tm}(n_{t-1} - n_{tm} + \rho_t^2 n_{tu}) + (1-\rho_t^2)n_{tu}(n_{t-1} - n_{tm})} . \quad (24)$$

Substituting the proper W_t', W_t'' and W_t from (22), (23) and (24) in (14), (15) and (16), respectively, we find that the minimum variances of $\hat{\bar{Y}}_t', \hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$ are

$$V(\hat{\bar{Y}}_t') = S_t^2\left(\frac{1}{n_t} - \frac{1}{N}\right) , \quad (25)$$

$$V(\hat{\bar{Y}}_t'') = S_t^2\left(\frac{(1-\rho_t^2)n_{t-1} + \rho_t^2 n_{tm}}{n_{t-1}n_{tm} + n_{tu}\left((1-\rho_t^2)n_{t-1} + \rho_t^2 n_{tm}\right)} - \frac{1}{N}\right) , \quad (26)$$

$$V(\hat{\bar{Y}}_t) = S_t^2\left(\frac{1}{n_{t-1} + n_{tu}} - \frac{1}{N}\right) + \frac{S_t^2(1-\rho_t^2)(n_{t-1} - n_{tm})}{(n_{t-1} + n_{tu})^2}\left(\frac{n_{t-1}(n_{t-1} - n_{tm} + \rho_t^2 n_{tu}) + (1-\rho_t^2)n_{tu}(n_{t-1} - n_{tm})}{n_{tm}(n_{t-1} - n_{tm} + \rho_t^2 n_{tu}) + (1-\rho_t^2)n_{tu}(n_{t-1} - n_{tm})}\right) , \quad (27)$$

respectively.

3. Comparison of the Estimators

3.1 Consistency

When $n_t = N$, we have

$$\hat{\bar{Y}}_t' = W_t'\bar{y}_{tu} + (1 - W_t')\bar{y}_{tm} ,$$

$$\hat{\bar{Y}}_t'' = W_t''\bar{y}_{tu} + (1 - W_t'')\bar{y}_{tm} ,$$

$$\hat{\bar{Y}}_t = \bar{Y}_t .$$

So $\hat{\bar{Y}}_t$ is always a consistent estimator of $\bar{Y}_t$, whereas $\hat{\bar{Y}}_t'$ and $\hat{\bar{Y}}_t''$ may be inconsistent if W_t' for $\hat{\bar{Y}}_t'$ and W_t'' for $\hat{\bar{Y}}_t''$ are not properly chosen.

3.2 Bias

From (11), (12) and (13), the biases of $\hat{\bar{Y}}_t', \hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$ are

$$B(\hat{\bar{Y}}_t') = 0 , \quad (28)$$

$$B(\hat{\bar{Y}}_t'') = (1-W_t'')C(b_t, \bar{y}_{t-1} - \bar{y}_{t-1,m^*}), \tag{29}$$

$$B(\hat{\bar{Y}}_t) = (1-W_t)\frac{n_{t-1}}{n_{t-1}+n_{tu}}C(b_t, \bar{y}_{t-1} - \bar{y}_{t-1,m^*}), \tag{30}$$

respectively. So $\hat{\bar{Y}}_t'$ is always an unbiased estimator of $\bar{Y}_t$, whereas $\hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$ will be unbiased if b_t is a pre-assigned constant or $n_{tm} = n_{t-1}$.

When $W_t'' = W_t$, it is obvious that $\left|B(\hat{\bar{Y}}_t)\right| \le \left|B(\hat{\bar{Y}}_t'')\right|$. When the proper W_t'' and W_t from (23) and (24) are used, the biases of $\hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$ become

$$B(\hat{\bar{Y}}_t'') = \frac{n_{t-1}n_{tm}}{n_{t-1}n_{tm} + n_{tu}\left((1-\rho_t^2)n_{t-1} + \rho_t^2 n_{tm}\right)}C(b_t, \bar{y}_{t-1} - \bar{y}_{t-1,m^*}), \tag{31}$$

$$B(\hat{\bar{Y}}_t) = \frac{n_{t-1}n_{tm}}{n_{t-1}n_{tm} + n_{tu}\left((1-\rho_t^2)n_{t-1} + \rho_t^2 n_{tm} + a\right)}C(b_t, \bar{y}_{t-1} - \bar{y}_{t-1,m^*}), \tag{32}$$

respectively, where $a = \dfrac{(1-\rho_t^2)(n_{t-1}-n_{tm})\left(n_{tm} + (1-\rho_t^2)n_{tu}\right)}{n_{t-1} - n_{tm} + \rho_t^2 n_{tu}}$. Since $a \ge 0$, $\left|B(\hat{\bar{Y}}_t)\right| \le \left|B(\hat{\bar{Y}}_t'')\right|$.

3.3 Variance

When $W_t' = W_t'' = W_t = \varphi_t$, the variances of $\hat{\bar{Y}}_t', \hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$ may be written as

$$V(\hat{\bar{Y}}_t') = S_t^2\left(\frac{\varphi_t^2}{n_{tu}} + \frac{(1-\varphi_t)^2}{n_{tm}} - \frac{1}{N}\right), \tag{33}$$

$$V(\hat{\bar{Y}}_t'') = S_t^2\left(\frac{\varphi_t^2}{n_{tu}} + \frac{(1-\varphi_t)^2}{n_{tm}} - \frac{1}{N}\right) - b, \tag{34}$$

$$V(\hat{\bar{Y}}_t) = S_t^2\left(\frac{\varphi_t^2}{n_{tu}} + \frac{(1-\varphi_t)^2}{n_{tm}} - \frac{1}{N}\right) - c, \tag{35}$$

where $b = \dfrac{S_t^2(1-\varphi_t)^2\rho_t^2(n_{t-1}-n_{tm})}{n_{t-1}n_{tm}}$ and

$$c = \frac{S_t^2}{n_{tu}n_{tm}(n_{t-1}+n_{tu})^2}\Big(\left(\varphi_t n_{tm} - (1-\varphi_t)n_{tu}\right)^2(2n_{t-1} + n_{tu} - n_{tm}) + 2\varphi_t(1-\varphi_t)\rho_t^2 n_{tu}n_{tm}(n_{t-1}-n_{tm})$$
$$+ (1-\varphi_t)^2\rho_t^2 n_{t-1}n_{tu}(n_{t-1}-n_{tm})\Big).$$

Since $b \ge 0$ and $c \ge 0$, $V(\hat{\bar{Y}}_t'') \le V(\hat{\bar{Y}}_t')$ and $V(\hat{\bar{Y}}_t) \le V(\hat{\bar{Y}}_t')$. To compare $V(\hat{\bar{Y}}_t)$ with $V(\hat{\bar{Y}}_t'')$, we write

$$V(\hat{\bar{Y}}_t) = S_t^2\left(\frac{\varphi_t^2}{n_{tu}} + \frac{(1-\varphi_t)^2}{n_{tm}} - \frac{1}{N}\right) - b - \frac{S_t^2}{n_{t-1}n_{tu}n_{tm}(n_{t-1}+n_{tu})^2}\left(\varphi_t^2 d - 2\varphi_t(1-\varphi_t)e + (1-\varphi_t)^2 f\right), \tag{36}$$

where $d = n_{t-1}n_{tm}^2(2n_{t-1} + n_{tu} - n_{tm})$,

$e = n_{t-1}n_{tu}n_{tm}\left(2n_{t-1} + n_{tu} - n_{tm} - \rho_t^2(n_{t-1}-n_{tm})\right)$ and

$f = n_{tu}^2\left(n_{t-1}(2n_{t-1} + n_{tu} - n_{tm}) - \rho_t^2(n_{t-1}-n_{tm})(2n_{t-1}+n_{tu})\right)$.

We find that, $V(\hat{\bar{Y}}_t) \le V(\hat{\bar{Y}}_t'')$ if $0 \le \varphi_t \le \dfrac{e + f - \sqrt{e^2 - df}}{d + 2e + f}$ or $\dfrac{e + f + \sqrt{e^2 - df}}{d + 2e + f} \ge \varphi_t \ge 1$. Note that, when $\rho_t = 0$

or $n_{tm} = n_{t-1}$ we have $V(\hat{\bar{Y}}_t) \le V(\hat{\bar{Y}}_t'')$ for all φ_t.

When the proper W_t', W_t'' and W_t from (22), (23) and (24) are used, the variances of $\hat{\bar{Y}}_t', \hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$ may be written as

$$V(\hat{\bar{Y}}_t') = S_t^2 \left(\frac{1}{n_{tm} + n_{tu}} - \frac{1}{N} \right), \tag{37}$$

$$V(\hat{\bar{Y}}_t'') = S_t^2 \left(\frac{1}{g n_{tm} + n_{tu}} - \frac{1}{N} \right), \tag{38}$$

$$V(\hat{\bar{Y}}_t) = S_t^2 \left(\frac{1}{n_{tm} + n_{tu}} - \frac{1}{N} \right) - h, \tag{39}$$

where $g = \dfrac{n_{t-1}}{n_{t-1} - \rho_t^2 (n_{t-1} - n_{tm})}$ and

$$h = \frac{S_t^2 \rho_t^2 n_{tm} (n_{t-1} - n_{tm})}{n_t (n_{t-1} + n_{tu})^2} \left(\frac{(n_{t-1} + n_{tu})(n_{t-1} - n_{tm} + \rho_t^2 n_{tu}) + (1-\rho_t^2) n_{tu} (n_{t-1} - n_{tm})}{n_{tm} (n_{t-1} - n_{tm} + \rho_t^2 n_{tu}) + (1-\rho_t^2) n_{tu} (n_{t-1} - n_{tm})} \right).$$

Since $g \ge 1$ and $h \ge 0$, $V(\hat{\bar{Y}}_t'') \le V(\hat{\bar{Y}}_t')$ and $V(\hat{\bar{Y}}_t) \le V(\hat{\bar{Y}}_t')$. To compare $V(\hat{\bar{Y}}_t)$ with $V(\hat{\bar{Y}}_t'')$, we write

$$V(\hat{\bar{Y}}_t'') = S_t^2 \left(\frac{1}{n_{t-1} + n_{tu}} - \frac{1}{N} \right) + \frac{S_t^2 (1-\rho_t^2)(n_{t-1} - n_{tm})}{(n_{t-1} + n_{tu})^2} \left(\frac{pq - pr}{q(q-r)} \right), \tag{40}$$

$$V(\hat{\bar{Y}}_t) = S_t^2 \left(\frac{1}{n_{t-1} + n_{tu}} - \frac{1}{N} \right) + \frac{S_t^2 (1-\rho_t^2)(n_{t-1} - n_{tm})}{(n_{t-1} + n_{tu})^2} \left(\frac{pq - qs}{q(q-r)} \right), \tag{41}$$

where $p = n_{t-1}(n_{t-1} + n_{tu})$,

$q = n_{t-1} n_{tm} + n_{tu} \left((1-\rho_t^2) n_{t-1} + \rho_t^2 n_{tm} \right)$,

$r = n_{tm} \left(n_{tm} + (1-\rho_t^2) n_{tu} \right)$ and

$s = n_{tm} \left(n_{t-1} + (1-\rho_t^2) n_{tu} \right)$.

Since $pr \ge qs$, $V(\hat{\bar{Y}}_t) \ge V(\hat{\bar{Y}}_t'')$. That is, when the proper weights which minimize the variances of the estimators are used we have $V(\hat{\bar{Y}}_t'') \le V(\hat{\bar{Y}}_t) \le V(\hat{\bar{Y}}_t')$.

4. Simulation Results

In order to compare the mean square errors of $\hat{\bar{Y}}_t', \hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$, the Monte Carlo simulation results are carried out. The data of the total turnover in each quarter of every enumerable sample establishment which has 1-100 workers from the Thailand Quarterly Retail Survey 2005, collected by the National Statistical Office of Thailand, is used as the population data.

From the data, we have $N = 1,096$, $\bar{Y}_1 = 1,582,365.2491$ baht, $\bar{Y}_2 = 1,661,515.5748$ baht, $\bar{Y}_3 = 1,716,284.2573$ baht, $\bar{Y}_4 = 1,842,827.1661$ baht, $S_1^2 = 4.0339 \times 10^{13}$ baht^2, $S_2^2 = 4.2046 \times 10^{13}$ baht^2, $S_3^2 = 4.6919 \times 10^{13}$ baht^2, $S_4^2 = 6.9422 \times 10^{13}$ baht^2, $\rho_2 = 0.9470$, $\rho_3 = 0.9892$ and $\rho_4 = 0.9832$.

We assume that $n_1 = n_2 = n_3 = n_4 = n$ and $n_{2m} = n_{3m} = n_{4m} = n_m$. n_m is varied, where $n = 100$ and $n_m = 25\%n,\ 50\%n,\ 75\%n$. The samples are selected by the simulation with 10,000 runs. The formula for computing the mean square errors of $\hat{\bar{Y}}_t', \hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$, $t = 2,3,4$, is

$$MSE(\tilde{\bar{Y}}_t) = \frac{\sum_{r=1}^{10,000} \left((\tilde{\bar{Y}}_t)_r - \bar{Y}_t \right)^2}{10,000}, \qquad (42)$$

where $\tilde{\bar{Y}}_t$ denotes $\hat{\bar{Y}}_t'$, $\hat{\bar{Y}}_t''$ or $\hat{\bar{Y}}_t$.

$\hat{\bar{Y}}_t'$, $\hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$ are considered for all possible weights, $0 \le W_t', W_t'', W_t \le 1$. Let φ_t denote the weights W_t', W_t'' and W_t, $t = 2,3,4$. Fig. 1 shows the mean square errors of $\hat{\bar{Y}}_t'$, $\hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$ by φ_t, $t = 2,3,4$. The unit of the mean square errors is $\times 10^{11}$ baht^2.

It is certain that the mean square error of each estimator depends on φ_t. For each φ_t, the mean square error of $\hat{\bar{Y}}_t'$ is never less than the mean square errors of the others. The minimum mean square errors of $\hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$ are close together, and among all possible φ_t the range of the mean square errors of $\hat{\bar{Y}}_t$ is smaller than that of $\hat{\bar{Y}}_t''$. The percentage of the sample units to match affects the mean square errors of $\hat{\bar{Y}}_t'$ and $\hat{\bar{Y}}_t''$, but hardly affects the mean square error of $\hat{\bar{Y}}_t$. Using improper φ_t, the mean square errors of $\hat{\bar{Y}}_t'$ and $\hat{\bar{Y}}_t''$ may be very high, whereas the mean square error of $\hat{\bar{Y}}_t$ is slightly different from the minimum.

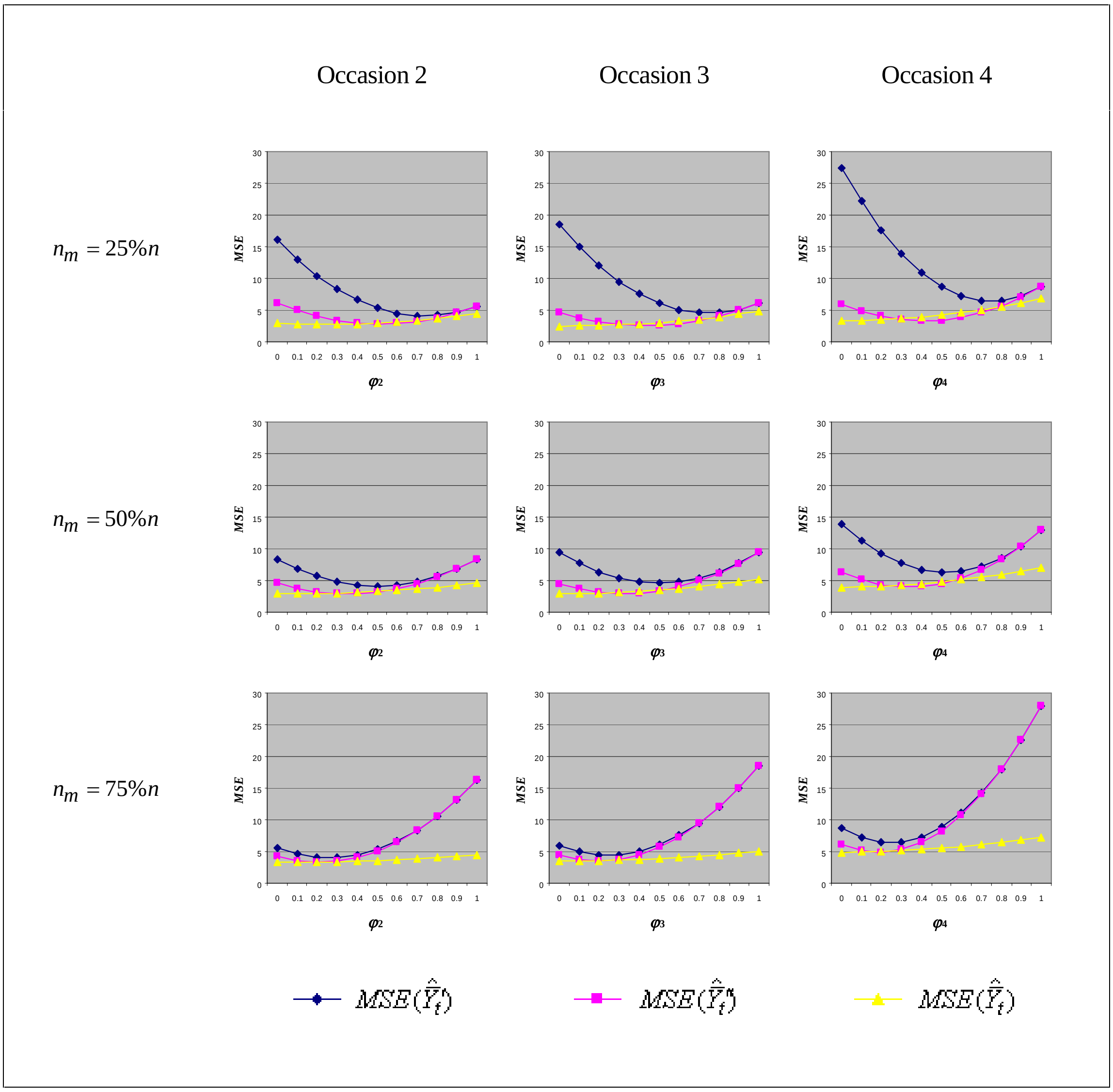

Fig. 1 The mean square errors of $\hat{\bar{Y}}_t'$, $\hat{\bar{Y}}_t''$ and $\hat{\bar{Y}}_t$ by φ_t, $t = 2,3,4$

5. Conclusions

An alternative estimator in finite population sampling on successive occasions is proposed. We compare it with an unbiased estimator and a conventional estimator. The alternative estimator is always consistent, whereas the others are consistent for some weights used in the estimation. The conventional estimator and the alternative estimator are biased if the regression coefficient used in the estimation is not a pre-assigned constant, where the absolute value of the bias of the alternative estimator is less than that of the other. The variance of the unbiased estimator is more than the variances of the others. For an interval of weights used in the estimation, the variance of the alternative estimator is not less than that of the conventional estimator. Using improper weights, the variance of the unbiased estimator and the mean square error of the conventional estimator may be very high, whereas the mean square error of the alternative estimator is slightly different from the minimum.

References

[1] Cochran, W. G.; Sampling Techniques, 3rd ed., New York: Wiley, 1977.

[2] Jessen, R. J.; Statistical Investigation of a Sample Survey for Obtaining Farm Facts, Iowa State College of Agriculture and Mechanic Arts Research Bulletin, No. 304, pp 54-59, 1942.

[3] Patterson, H. D.; Sampling on Successive Occasions with Partial Replacement of Units, Journal of the Royal Statistical Society, 12 (Series B), pp 241-255, 1950.

[4] Yates, F.; Sampling Methods for Censuses and Surveys, 3rd ed., New York: Hafner, 1960.