# TREATMENT OF UNIT NON-RESPONSE IN TWO-STAGE SAMPLING WITH PARTIAL REPLACEMENT OF UNITS

Fabian C. Okafor (*)

## 1. INTRODUCTION

Many authors have worked on sampling on successive occasions; among them are Jessen (1942), Singh (1968), Kathuria, *et al.* (1971), Okafor (1988) to mention but a few. All of the above researchers assumed total response from all the sample units on all the occasions. Singh, et al (1974) using Bartholomew (1961) method of treating unit non-response proposed an estimator of the population mean for sampling on two successive occasions. Okafor and Lee (2000) applied Hansen and Hurwitz (1946) technique for treating unit non-response to double sampling for ratio and regression estimation. Okafor (2001) discussed the treatment of unit non-response in successive sampling over two occasions for element sampling.

In this paper we have attempted the extension of the treatment of unit non-response to two-stage sampling with partial replacement of units at the second stage using Hansen and Hurwitz (1946) technique. We have proposed three different estimators for the population total when sampling with partial replacement of units and with equal probability sampling at both stages. These estimators have been compared empirically using actual data from a survey on arable farming conducted by the Ministry of Agriculture, Botswana.

## 2. SAMPLING SCHEME

*First Occasion:*

Select $n$ first stage units (fsu's) from the population of $N$ fsu's by simple random sampling (srs). Within the $i^{th}$ selected fsu, select a sample of $m_i$ second stage units (ssu's) from the population of $M_i$ ssu's again by srs. Information on the study variable $y$ is sought from the sample ssu's. Suppose $m_{1i}$ units respond at the first attempt and $m_{2i}$ of the $m_i$ ssu's not respond. Let a simple random subsample of $s_{1i}$ $(= k_i m_{2i};\ 0 < k_i \leq 1)$ units from the $m_{2i}$ ssu's be revisited. We shall assume that all the $s_{1i}$ ssu's gave information on $y$ at this second attempt.

(*) This study was carried out while in University of Botswana, Gaborone, Botswana.

*Second Occasion:*

All the $n$ fsu's selected at the first occasion are retained for use at the second occasion. Select by srs $\lambda m_i$ $(0<\lambda<1)$ ssu's from the $m_i$ ssu's already selected at the first occasion. Suppose $m'_{1i}$ ssu's out of the $\lambda m_i$ ssu's respond at the first attempt and $m'_{2i}$ fail to respond. Let a simple random subsample of $s'_{1i}$ $(=k_i m'_{2i})$ units be revisited. We again assume a complete response of all the $s'_{1i}$ units at this second attempt.

In the $i^{th}$ sample fsu, select a fresh sample of $\mu m_i$ $(\lambda+\mu=1)$ ssu's from $M_i$ ssu's by srs. For these fresh $\mu m_i$ units, let $u_{1i}$ ssu's respond and $u_{2i}$ fail to supply the required information on the study variable. A simple random subsample of size $s_{2i}$ $(=k_i u_{2i})$ from $u_{2i}$ ssu's is revisited. A complete response is again assumed this time around.

## 3. ESTIMATION OF THE POPULATION TOTAL

For simplicity, we make the following assumptions:
$\lambda m_i$ matched ssu's at the second occasion form a subset of the first occasion, $m_{1i}+s_{1i}$, responding units.

Sampling is from infinite population at both stages. $M_i$ and the population variability remain constant at both occasions.

### 3.1. *Notations*

Let $x$ ($y$) be the values of the study variate at the first (second) occasion.

$\overline{x}^*_{m_i}=(m_{1i}\overline{x}_{1i}+m_{2i}\overline{x}_{s_{1i}})/m_i$ is the first occasion mean of the $i^{th}$ fsu adjusted for non- response based on $m_i$ ssu's.

$\overline{x}_{1i}$ is the first occasion sample mean of the $i^{th}$ fsu based on the $m_{1i}$ ssu's.

$\overline{x}_{s_{1i}}$ is the first occasion sample mean of the $i^{th}$ fsu based on the $s_{1i}$ responding unit at the second attempt.

$\overline{z}^*_i=(m'_{1i}\overline{z}'_{1i}+m'_{2i}\overline{z}'_{s'_{1i}})/\lambda m_i$ ; (z = $x$, $y$) is the sample mean adjusted for non-response based on the matched sample of ssu's.

$\overline{z}'_{1i}$ is the sample mean for the matched sample of ssu's based on the $m'_{1i}$ respondents at the first attempt.

$\overline{z}'_{s'_{1i}}$ is the sample mean for the matched sample of ssu's in the $i^{th}$ fsu based on the $s'_{1i}$ respondents at the second attempt

$\overline{y}^*_{u_i}=(u_{1i}\overline{y}''_{1i}+u_{2i}\overline{y}''_{s_{2i}})/\mu m_i$ is the second occasion sample mean for the unmatched sample ssu's of the $i^{th}$ fsu adjusted for non-response.

$\overline{y}''_{1i}$ is the second occasion sample mean for the unmatched sample ssu's in the $i^{th}$ fsu based on the $u_{1i}$ respondents at the first attempt.

$\overline{y}''_{s_{2i}}$ is the second occasion sample mean for the unmatched sample ssu's in the $i^{th}$ fsu based on the $s_{2i}$ respondents at the second attempt.

3.2. *Estimators of the population total*

Let us consider the first estimator of the form:

$$T_1 = \frac{N}{n}\sum_{i=1}^{n} M_i[\theta_i\{\overline{y}_i^* - \beta_i^*(\overline{x}_i^* - \overline{x}_{m_i}^*)\} + (1-\theta_i)\overline{y}_{u_i}^*] \tag{3.1}$$

$\beta_i^*$ is a known constant of proportionality in the $i^{th}$ fsu

$\theta_i$ is the weight, in the $i^{th}$ fsu, chosen so as to make the variance of $T_1$ a minimum. The estimator $T_1$ is unbiased (see appendix for proof).

In (3.1) the first term inside the square brackets and enclosed in gothic brackets is a double sampling regression estimator, since the first occasion observation, $x$ is used as an auxiliary variable to improve the current estimate of the population mean in each fsu. This regression estimator is suitably weighted with the basic estimator, $\overline{y}_{u_i}^*$ obtained by using the fresh sample selected on the second occasion to form one estimator of the population mean within each fsu. The same procedure is adopted in obtaining the other remaining two estimators.

It may be cumbersome to compute the weight $\theta_i$, especially when the number of sample fsu is large. Therefore, it may be advisable to have a common weight for all the fsu's. Hence the estimator

$$T_2 = \theta\frac{N}{n}\sum_{i=1}^{n} M_i\{\overline{y}_i^* - \beta_i^*(\overline{x}_i^* - \overline{x}_{m_i}^*)\} + (1-\theta)\frac{N}{n}\sum_{i=1}^{n} M_i\overline{y}_{u_i}^* \tag{3.2}$$

A third estimator of the population total is obtained by using a common weight and a common regression coefficient, $\beta_c$. This estimator is of the form

$$T_3 = \phi\frac{N}{n}\sum_{i=1}^{n} M_i[\overline{y}_i^* - \beta_c(\overline{x}_i^* - \overline{x}_{m_i}^*)] + (1-\phi)\frac{N}{n}\sum_{i=1}^{n} M_i\overline{y}_{u_i}^* = \phi T_m + (1-\phi)T_u \tag{3.3}$$

3.3. *Variance of the proposed estimators*

To derive the variance of $T_1$, we adopt the conditional approach and follow the procedure used in Hansen and Hurwitz (1946) estimator. The derivation of the variance is given in the appendix.

$$V(T_1)=\frac{N^2}{n}S_b^2+\frac{N}{n}\sum_{i=1}^{N}M_i^2\left[\left\{\theta_i^2\frac{(1-\mu\rho_i^2)}{\lambda m_i}+\frac{(1-\theta_i)^2}{\mu m_i}\right\}S_i^2\right.$$
$$\left.+V_{2i}\left\{\frac{\theta_i^2(1+\beta_i^{*2}-2\beta_i^*\rho_{2i})}{\lambda m_i}\right\}+\frac{(1-\theta_i)^2}{\mu m_i}\right] \qquad (3.4)$$

where,

$$V_{2i}=W_{2i}(1-k_i)S_{2i}^2\big/k_i$$

$S_b^2$ is the population variability between the fsu's.

$S_i^2$ is the population variability within the $i^{th}$ fsu.

$S_{2i}^2$ is the population variability of the non-respondents within the $i^{th}$ fsu.

$\rho_i$ is the population correlation coefficient between the first and second occasions for the $i^{th}$ fsu.

$\rho_{2i}$ is the within fsu population correlation coefficient between the first and second occasions for the non-respondents.

$W_{2i}$ is the population proportion of the non-respondents.

By differentiating V($T_1$) with respect to $\theta_i$ and solving the derivative, we have the optimum value of $\theta_i$ as

$$\theta_{0i}=\lambda V_i\big/[(1-\mu^2\rho_i^2)S_i^2+V_{2i}\{1+\mu(\beta_i^{*2}-2\beta_i^*\rho_{2i})\}] \qquad (3.5)$$
$$V_i=S_i^2+V_{2i}$$

Substituting $\theta_{0i}$ in (3.5) and simplifying, the minimum value of V($T_1$) becomes

$$V_0(T_1)=\frac{N^2}{n}S_b^2+\frac{N}{n}\sum_{i=1}^{N}\frac{M_i^2}{\mu m_i}(1-\theta_{0i})V_i \qquad (3.6)$$

The optimum value of $\beta_i^*$ is $\beta_{0i}^*=(\rho_i S_i^2+\rho_{2i}V_{2i})\big/V_i$

We shall use the same procedure for obtaining the variance and the optimum weight for the estimator, $T_1$ to derive the variance of $T_2$ and its optimum weight.

Hence the optimum value of $\theta$ is

$$\theta_0=\lambda\sum_{i=1}^{N}\frac{M_i^2}{m_i}V_i\bigg/\sum_{i=1}^{N}\frac{M_i^2}{m_i}[S_i^2(1-\mu^2\rho_i^2)+V_{2i}\{1+\mu(\beta_i^{*2}-2\beta_i^*\rho_{2i})\}]$$

While the minimum variance of $T_2$ is

$$V_o(T_2)=\frac{N^2}{n}S_b^2+(1-\theta_o)\frac{N}{n}\sum_{i=1}^{N}\frac{M_i^2}{\mu m_i}V_i \qquad (3.7)$$

Finally by minimizing the variance of $T_3$ with respect to $\phi$, the optimum value of $\phi$ is

$$\phi_o = \{V(T_u) - Cov(T_m, T_u)\} / \{V(T_m) + V(T_u) - 2Cov(T_m, T_u)\} \tag{3.8}$$

While the minimum variance is

$$V_o(T_3) = \{V(T_u)V(T_m) - \text{Cov}^2(T_m, T_u)\} / \{V(T_m) + V(T_u) - 2Cov(T_m, T_u)\} \tag{3.9}$$

Where,

$$Cov(T_m, T_u) = \frac{N^2}{n} S_b^2 \,;\; V(T_u) = \frac{N^2}{n} S_b^2 + \frac{N}{n} \sum_{i=1}^{N} \frac{M_i^2}{\mu m_i} V_i \tag{3.10}$$

$$V(T_m) = \frac{N^2}{n} S_b^2 + \frac{N}{n} \sum_{i=1}^{N} \frac{M_i^2}{\lambda m_i} V_i + \beta_c^2 \frac{N}{n} \sum_{i=1}^{N} \frac{M_i^2}{\lambda m_i} \{\mu S_i^2 + V_{2i}\}$$

$$-2\beta_c \frac{N}{n} \sum_{i=1}^{N} \frac{M_i^2}{\lambda m_i} \{\mu \rho_i S_i^2 + \rho_{2i} V_{2i}\} \tag{3.11}$$

The optimum value of $\beta_c$ obtained by minimizing the variance of $T_m$ , V($T_m$) is

$$\beta_c = \left[ \sum_{i=1}^{N} \frac{M_i^2}{m_i} \{\mu \rho_i S_i^2 + \rho_{2i} V_{2i}\} \right] \Bigg/ \left[ \sum_{i=1}^{N} \frac{M_i^2}{m_i} \{\mu S_i^2 + V_{2i}\} \right]$$

## 4. COMPARISON OF THE ESTIMATORS

### 4.1. *Theoretical comparison*

To gain insight into the performance of the proposed estimators and for simplicity, we shall compare only the estimator $T_2$ with the estimator, $T_0$ obtained when there is no partial matching of units.

$$T_0 = \frac{N}{n} \sum_{i=1}^{n} M_i \bar{y}_{m_i}^* \tag{4.1}$$

with variance

$$V(T_0) = \frac{N^2}{n} S_b^2 + \frac{N}{n} \sum_{i=1}^{N} \frac{M_i^2}{m_i} V_i \tag{4.2}$$

For the purpose of comparison, we shall make the following assumptions:

$k_i = k$, $W_{2i} = W_2$; $\beta_i^* = \rho_i = \rho$ and $\rho_{2i} = \rho$ for all first stage units.

Based on these assumptions

$$V(T_0) = \frac{N^2}{n}\left[ S_b^2 + S_w^2 + \frac{W_2(1-k)}{k} S_{2w}^2 \right] \tag{4.3}$$

$$S_w^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{M_i^2}{m_i}S_i^2 \; ; \; S_{2w}^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{M_i^2}{m_i}S_{2i}^2$$

The efficiencies of $T_2$ with respect to $T_o$ have been calculated for different values of $\delta = S_w^2 / S_b^2$, $\delta_2 = S_{2w}^2 / S_b^2$, $\rho$, $\rho_2$, $k$ and $W_2$. The results are presented in tables 1 and 2 below. We have to note that since variances of $T_0$ and $T_2$ involve $W_2$ and $k$, the individual variances will be high if $W_2$ is high or $k$ is small. So the efficiency of $T_2$ with respect to $T_o$ is based on the given values of $W_2$ and $k$.

From tables 1 and 2, we observe that $T_2$ is not necessarily more efficient than $T_o$ when $k$ = 0.4 because here the efficiency of $T_2$ over $T_o$ is less than one when $\delta_2$ =0.5.

This may be partly due to the assumptions used in the comparison and partly due to the nature of the estimator. But in most cases there is much to be gained in using the estimator $T_2$ in place of $T_o$ For instance under the given assumptions, there is a gain in efficiency of $T_2$ over $T_o$ when $k$= 0.1. The efficiency is higher for $W_2$ =0.7 than for $W_2$ =0.3. For a given value of $\rho$, the gain inefficiency increases for increasing value of $\rho_2$. Also for a given value of $\rho_2$ other than 0.1, the efficiency increases for increasing value of $\rho$. Further more, the efficiency of $T_2$ over $T_o$ increases for increasing value of $\delta_2$; but decreases for increasing value of $\delta$.

TABLE 1

*Efficiency of $T_2$ with respect to $T_o$ for $W_2$ = 0.3*

| δ | | | 0.5 | | | | | | 1.5 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\varrho$ | 0.1 | | 0.5 | | 0.9 | | 0.1 | | 0.5 | | 0.9 | |
| $\delta_2$ | $k$ | $\rho_2$ | $\mu$ 0.2 | 0.8 | $\mu$ 0.2 | 0.8 | $\mu$ 0.2 | 0.8 | $\mu$ 0.2 | 0.8 | $\mu$ 0.2 | 0.8 | $\mu$ 0.2 | 0.8 |
| 0.5 | 0.1 | 0.1 | 1.26 | 1.26 | 1.21 | 1.24 | 1.09 | 1.16 | 0.96 | 0.96 | 0.95 | 1.01 | 0.90 | 1.07 |
| | | 0.5 | 1.29 | 1.25 | 1.37 | 1.23 | 1.31 | 1.19 | 0.98 | 0.94 | 1.03 | 0.92 | 1.04 | 1.04 |
| | | 0.9 | 1.33 | 1.25 | 1.62 | 1.21 | 1.82 | 1.33 | 1.00 | 0.92 | 1.16 | 0.77 | 1.31 | 0.92 |
| | 0.4 | 0.1 | 0.80 | 0.79 | 0.80 | 0.85 | 0.81 | 0.97 | 0.36 | 0.36 | 0.37 | 0.42 | 0.41 | 0.67 |
| | | 0.5 | 0.80 | 0.78 | 0.82 | 0.77 | 0.85 | 0.91 | 0.36 | 0.34 | 0.37 | 0.34 | 0.40 | 0.54 |
| | | 0.9 | 0.80 | 0.76 | 0.85 | 0.63 | 0.91 | 0.73 | 0.35 | 0.32 | 0.36 | 0.23 | 0.39 | 0.29 |
| 1.5 | 0.1 | 0.1 | 1.57 | 1.56 | 1.46 | 1.48 | 1.22 | 1.26 | 1.37 | 1.36 | 1.29 | 1.34 | 1.12 | 1.22 |
| | | 0.5 | 1.64 | 1.57 | 1.80 | 1.53 | 1.64 | 1.34 | 1.42 | 1.36 | 1.54 | 1.32 | 1.45 | 1.26 |
| | | 0.9 | 1.71 | 1.58 | 2.49 | 1.68 | 3.07 | 1.83 | 1.47 | 1.35 | 2.02 | 1.29 | 2.44 | 1.49 |
| | 0.4 | 0.1 | 1.08 | 1.07 | 1.05 | 1.09 | 0.99 | 1.09 | 0.71 | 0.70 | 0.71 | 0.78 | 0.72 | 0.95 |
| | | 0.5 | 1.09 | 1.06 | 1.14 | 1.05 | 1.13 | 1.09 | 0.71 | 0.68 | 0.74 | 0.67 | 0.77 | 0.86 |
| | | 0.9 | 1.11 | 1.05 | 1.27 | 0.96 | 1.38 | 1.08 | 0.72 | 0.66 | 0.78 | 0.51 | 0.86 | 0.63 |

TABLE 2

*Efficiency of $T_2$ with respect to $T_o$ for $W_2 = 0.7$*

| $\delta$ | | | 0.5 | | | | | | 1.5 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\varrho$ | 0.1 | | 0.5 | | 0.9 | | 0.1 | | 0.5 | | 0.9 | |
| $\delta_2$ | $k$ | $\rho_2$ | $\mu$ 0.2 | 0.8 | $\mu$ 0.2 | 0.8 | $\mu$ 0.2 | 0.8 | $\mu$ 0.2 | 0.8 | $\mu$ 0.2 | 0.8 | $\mu$ 0.2 | 0.8 |
| 0.5 | 0.1 | 0.1 | 1.50 | 1.49 | 1.40 | 1.43 | 1.19 | 1.24 | 1.28 | 1.27 | 1.22 | 1.27 | 1.07 | 1.19 |
| | | 0.5 | 1.56 | 1.50 | 1.70 | 1.46 | 1.57 | 1.30 | 1.32 | 1.26 | 1.43 | 1.24 | 1.36 | 1.22 |
| | | 0.9 | 1.62 | 1.50 | 2.26 | 1.57 | 2.71 | 1.71 | 1.36 | 1.25 | 1.79 | 1.17 | 2.12 | 1.35 |
| | 0.4 | 0.1 | 1.01 | 1.01 | 1.00 | 1.04 | 0.95 | 1.06 | 0.62 | 0.62 | 0.63 | 0.69 | 0.65 | 0.89 |
| | | 0.5 | 1.03 | 1.00 | 1.07 | 0.99 | 1.06 | 1.05 | 0.62 | 0.59 | 0.64 | 0.58 | 0.68 | 0.79 |
| | | 0.9 | 1.04 | 0.98 | 1.16 | 0.88 | 1.25 | 1.00 | 0.62 | 0.57 | 0.66 | 0.43 | 0.72 | 0.54 |
| 1.5 | 0.1 | 0.1 | 1.77 | 1.75 | 1.60 | 1.62 | 1.29 | 1.31 | 1.64 | 1.63 | 1.50 | 1.53 | 1.23 | 1.29 |
| | | 0.5 | 1.86 | 1.77 | 2.10 | 1.72 | 1.85 | 1.42 | 1.72 | 1.63 | 1.92 | 1.59 | 1.73 | 1.37 |
| | | 0.9 | 1.97 | 1.80 | 3.32 | 2.02 | 4.54 | 2.20 | 1.81 | 1.64 | 2.87 | 1.73 | 3.80 | 1.94 |
| | 0.4 | 0.1 | 1.30 | 1.30 | 1.25 | 1.27 | 1.11 | 1.17 | 1.02 | 1.02 | 1.00 | 1.06 | 0.93 | 1.10 |
| | | 0.5 | 1.34 | 1.30 | 1.43 | 1.27 | 1.36 | 1.21 | 1.04 | 1.00 | 1.10 | 0.98 | 1.10 | 1.08 |
| | | 0.9 | 1.38 | 1.29 | 1.73 | 1.27 | 1.96 | 1.40 | 1.06 | 0.98 | 1.26 | 0.84 | 1.43 | 0.99 |

We have to be extremely cautious in interpreting these efficiencies since the fact that the efficiency is higher for $k$= 0.1 or $W_2$ =0.7 does not mean that the estimate obtained from the estimator $T_2$ when $k$ is small is more precise than the one obtained when $k$ is large, as seen from table 5. The reason for this increase in efficiency is that the rate of increase of the variance of $T_2$ is lower than the rate of increase of the variance of $T_o$ as $k$ or $W_2$ changes.

### 4.2. *Empirical comparison*

The data used for the empirical comparison are from a survey conducted in 2000 on arable farming in Botswana. The original design was a stratified two-stage design. The strata consisted of geographical districts. In each district a sample of agricultural extension areas (land area covered by an extension worker) was selected. Within each selected extension area, a sample of farmers from the list of registered farmers in the area was selected. Information was then collected on hectares of land owned and land planted in 1999/2000 agricultural season and past four seasons from the selected farmers. Information on other variables of interest was also obtained from the farmers.

For the purpose of our empirical comparison, we have used the sample data from only five districts (Ngwaketse West, South, Central and North and Barolong) as our population data. First occasion is the 1998/1999 planting season; while the second occasion is the 1999/2000 planting season. We shall use the districts as our first stage units and farmers within districts as the second stage units. Using these data we calculated the population parameters presented in tables 3 and 4 below. We have hypothetically defined our non-respondent categories using the sample data as follows: In Ngwaketse West and Central districts, we assumed that the non-respondent stratum is made up of those farmers owning land of more than or equal to 16 hectares; in Ngwaketse South, it is those having more than or equal to 14 hectares of land. Non-respondent stratum in Ngwaketse North and Barolong districts consist of those farmers with land equal to or more than 10 and 12 hectares respectively. The large non-respondent variances shown in table 4 is due to the fact that there is very wide disparity in the size of land owned by this category of units according to our categorizing procedure.

TABLE 3

*Population parameters (overall) for 1999/2000*

| *District* | *Total (Y)* | $M_i$ | $S_i^2$ | $\beta$ | $\varrho$ |
|---|---|---|---|---|---|
| Ngwaketse West | 281.0002 | 56 | 26.9434 | 0.621 | 0.669 |
| Ngwaketse South | 167.6194 | 36 | 40.0094 | 0.536 | 0.429 |
| Ngwaketse Central | 253.5008 | 28 | 68.5981 | 0.744 | 0.884 |
| Ngwaketse North | 99.9994 | 27 | 15.9648 | 1.422 | 0.743 |
| Barolong | 400.6005 | 37 | 1222.8939 | 1.083 | 0.961 |

TABLE 4

*Population parameters (non-respondents) for 1999/2000*

| *District* | *Response Category* | $W_2$ | $S_i^2$ | $\varrho$ |
|---|---|---|---|---|
| Ngwaketse West | Respondent | | 5.8144 | 0.422 |
| | Non-respondent | 0.43 | 37.5867 | 0.648 |
| Ngwaketse South | Respondent | | 7.9163 | -0.042 |
| | Non-respondent | 0.36 | 48.0831 | 0.288 |
| Ngwaketse Central | Respondent | | 5.3389 | 0.035 |
| | Non-respondent | 0.50 | 83.6914 | 0.880 |
| Ngwaketse North | Respondent | | 1.5638 | 0.190 |
| | Non-respondent | 0.44 | 22.6119 | 0.789 |
| Barolong | Respondent | | 12.8092 | 0.539 |
| | Non-respondent | 0.19 | 1165.2324 | 0.924 |

$S_b^2 = 13123.362$; $S_i^2$ is the population variability within each district.

We have chosen within fsu sampling fraction of 0.1 for the purpose of computing the sampling variances of the proposed estimators.

The variances of the proposed estimators for a matching fraction of $\lambda = 0.6$ are presented in table 5 below. We observe that the estimator $T_1$ is the best among all the other estimators. This estimator is obtained by using individual within fsu weights and regression coefficients.

TABLE 5

*Variances of the proposed estimators*

| | *k= 0.2* | | *k= 0.8* | |
|---|---|---|---|---|
| *Estimators* | *Variance* | *Efficiency* | *Variance* | *Efficiency* |
| $T_o$ | $1.05490\times10^6$ | 1.000 | $9.98329\times10^5$ | 1.000 |
| $T_1$ | $8.02867\times10^5$ | 1.314 | $7.63119\times10^5$ | 1.308 |
| $T_2$ | $8.03857\times10^5$ | 1.312 | $7.64353\times10^5$ | 1.306 |
| $T_3$ | $8.10901\times10^5$ | 1.301 | $7.68318\times10^5$ | 1.299 |

*Department of Statistics*
*University of Nigeria,*
*Nsukka*

FABIAN C. OKAFOR

APPENDIX

The estimator $T_1$ is unbiased.

*Proof*

$$E(T_1) = E_1 E_2 E_2 E_{s_2}(T_1) \tag{A.1}$$

$E_1$ = expectation over the first stage sample.
$E_2$ = conditional expectation over the second stage sample given the first stage.
$E'_2$ = conditional expectation over matched sample in the $i^{th}$ fsu.
$E_{s_2}$ = conditional expectation over the subsample of the second stage non-respondents in the $i^{th}$ fsu.

Taking expectations one by one

$$E_{s_2}(T_1) = \frac{N}{n}\sum_{i=1}^{n} M_i[\theta_i\{\bar{y}'_i - \beta_i^*(\bar{x}'_i - \bar{x}_i)\} + (1-\theta_i)\bar{y}_{u_i}] \tag{A.2}$$

where,

$$\bar{z}'_i = \frac{1}{\lambda m_i}\sum_{j=1} z_{ij}\ ,\ \bar{z}_i = \frac{1}{m_i}\sum_{j=1} z_{ij}\ ;\ z = x, y \text{ and } \bar{y}_{u_i} = \frac{1}{\mu m_i}\sum_{j=1} y_{ij}$$

$$E'_2 E_{s_2}(T_1) = \frac{N}{n}\sum_{i=1}^{n} M_i[\theta_i \bar{y}_i + (1-\theta_i)\bar{y}_{u_i}] \tag{A.3}$$

$$E_2 E'_2 E_{s_2}(T_1) = \frac{N}{n}\sum_{i=1}^{n} M_i \bar{Y}_i \tag{A.4}$$

Finally,

$$E_1 E_2 E'_2 E_{s_2}(T_1) = \sum_{i=1}^{N} Y_i = Y$$

This completes the proof.
The unbiasedness of other estimators is proved is a similar fashion.

Proof of variance of $T_1$

*Proof*

$$V(T_1) = V_1 E_2 E'_2 E_{s_2}(T_1) + E_1 V_2 E'_2 E_{s_2}(T_1) + E_1 E_2 V'_2 E_{s_2}(T_1) + E_1 E_2 E'_2 V_{s_2}(T_1)$$

$V_1, V_2, V_2'$ and $V_{s_2}$ are variances analogous to the expectations defined earlier.
From (A.4)

$$V_1 \mathrm{E}_2 \mathrm{E}_2' \mathrm{E}_{s_2}(T_1) = \frac{N}{n}\sum_{i=1}^{n} Y_i = \frac{N^2}{n}\sum_{i=1}^{N}(Y_i - \overline{Y})^2 \big/ (N-1) = \frac{N^2}{n} S_b^2 \tag{A.5}$$

From (A.3)

$$\mathrm{E}_1 V_2 \mathrm{E}_2' \mathrm{E}_{s_2}(T_1) = \mathrm{E}_1 V_2 \left[ \frac{N}{n}\sum_{i=1}^{n} M_i [\theta_i \overline{y}_i + (1-\theta_i)\overline{y}_{u_i}] \right]$$

$$= \frac{N}{n}\sum_{i=1}^{N} M_i^2 \left\{ \theta_i^2 \frac{S_i^2}{m_i} + (1-\theta_i)^2 \frac{S_i^2}{\mu m_i} \right\} \tag{A.6}$$

From (A.2)

$$\mathrm{E}_1 \mathrm{E}_2 V_2' \mathrm{E}_{s_2}(T_1) = \frac{N}{n}\sum_{i=1}^{N} M_i^2 \theta_i^2 \left\{ \frac{\mu}{\lambda m_i} S_i^2 (1-\rho_i^2) \right\} \tag{A.7}$$

Further more,

$$\mathrm{E}_1 \mathrm{E}_2 \mathrm{E}_2' V_{s_2}(T_1) = \frac{N}{n}\sum_{i=1}^{N} M_i^2 V_{2i} \left[ \frac{(1-\theta_i)^2}{\mu m_i} + \frac{\theta_i^2}{\lambda m_i}(1+\beta_i^{*2} - 2\beta_i^* \rho_{2i}) \right] \tag{A.8}$$

Now combining (A.5), (A.6), (A.7) and (A.8) the variance of $T_1$ is

$$V(T_1) = \frac{N^2}{n} S_b^2 + \frac{N}{n}\sum_{i=1}^{N} M_i^2 \left[ \left\{ \theta_i^2 \frac{(1-\mu\rho_i^2)}{\lambda m_i} + \frac{(1-\theta_i)^2}{\mu m_i} \right\} S_i^2 \right.$$

$$\left. + V_{2i} \left\{ \frac{\theta_i^2 (1+\beta_i^{*2} - 2\beta_i^* \rho_{2i})}{\lambda m_i} + \frac{(1-\theta_i)^2}{\mu m_i} \right\} \right] \tag{A.9}$$

## REFERENCES

BARTHOLOMEW, D.J. (1961): *A method of allowing not-at-home bias in sample surveys,* "Applied Statistics", 10, pp. 52-59.

HANSEN, M.H. and HURWITZ, W.N. (1946): *The problem of non-response in sample surveys,* "Journal of American Statistical Association",41, pp. 517-529.

JESSEN, R.J. (1942): *Statistical investigation of a sample survey for obtaining farm facts.* " Statistics Research Bulletin, 304, Iowa State College of Agriculture".

KATHURIA, O. P. and SINGH, D. (1971): *Relative efficiencies of some alternative procedures in two-stage sampling on successive occasions,* "Journal of Indian Society of Agricultural Statistics", 23, pp. 101-114.

OKAFOR, F.C. (1988): *Ratio-to-size estimators of mean per subunit in two-stage sampling over two occasions*, "Australian Journal of Statistics", 30, 3, pp. 367-378.

OKAFOR, F.C. (2001): *Treatment of non-response in successive sampling,* "Statistica", anno LXI, 2, pp. 195-204.

OKAFOR, F. C. and LEE, H. (2000): *Double sampling for ratio/regression estimation with subsampling for non-response*, "Survey Methodology, Canada", 26, 2, pp. 183-188.

SINGH, D. (1968): *Estimates in successive sampling using a multistage design*, "Journal of American Statistical Association", 63, pp. 99-112.

SINGH, D., SINGH, R. and SINGH, P. (1974): *A study of non-response in successive sampling.* "Journal Indian Society of Agricultural Statistics", 26, 2, pp. 37-41.

## RIASSUNTO

*Trattamento delle mancate risposte totali nel campionamento a due stadi con sostituzione parziale delle unità*

Vengono presentati tre stimatori per il totale di popolazione basati su un disegno campionario a due stadi con sostituzione parziale solo delle unità del secondo stadio. Gli stimatori proposti sono costruiti tenendo conto dell'ipotesi di presenza di mancata risposta totale, e l'aggiustamento per la mancata risposta è basato sulla tecnica proposta da Hansen e Hurwitz (1946). Da un'analisi empirica emerge che gli stimatori per il totale di popolazione che usano pesi individuali per le unità di primo stadio hanno migliori prestazioni degli stimatori che usano pesi comuni. Sul versante teorico viene dimostrato che maggiore è il rapporto fra la variabilità entro le unità di primo stadio relative ai non rispondenti e la variabilità fra le unità di primo stadio, maggiore è il guadagno che si ottiene, in termini di efficienza, usando gli stimatori proposti invece degli stimatori che si ottengono quando non c'è *matching* parziale delle unità.

## SUMMARY

*Treatment of unit non-response in two-stage sampling with partial replacement of units*

Three separate estimators, for the estimation of the population total, based on two-stage sampling design on two successive occasions with partial replacement of secondary stage units only have been presented. For these estimators it is assumed that there is unit non-response. Hence, Hansen and Hurwitz (1946) technique has been used to adjust for the non-response. Empirically, it has been found that the estimator that uses individual weights within first stage units perform better than the other estimators that use common weights in estimating the population total. Theoretically, it has also been shown that the larger the ratio of the within first stage unit variability of the non-respondents to the between first stage unit variability the higher the gain in efficiency of the proposed estimators over the estimator obtained when there is no partial matching of units.