ON RATIO ESTIMATION IN POSTSTRATIFIED SAMPLING OVER TWO OCCASIONS

A. C. ONYEKA

(Received 30 August 2002; Revision accepted 18 December 2002)

Abstract

Two estimators are proposed for the estimation of the second occasion population ratio, \( R_2 \), of two characters of study in poststratified sampling over two occasions. The estimators are proposed along the line of Tripathi and Sinha (1976) and Okafor (1985). One of the estimators, \( d_1 \), is a ratio-cum-product type estimator while the other, \( d_2 \), is a product-cum-ratio type estimator. Both estimators do not assume knowledge of the first occasion population ratio, \( R_1 \). Expressions for the optimum matching or replacement fractions of both estimators are obtained since the estimators are based on a partial replacement of sample units on the second occasion. Conditions under which one estimator is to be preferred to the other estimator are obtained for repeated samples of fixed sizes.

Keywords: Successive sampling, poststratified sampling, ratio estimation, matching or replacement fraction, repeated samples.

INTRODUCTION

Onyeaka (2002) extended the theory of ratio estimation in a one-time poststratified sampling (PSS) to PSS over successive occasions. He proposed two estimators, \( e_1 \) and \( e_2 \), for the estimation of the second occasion population ratio, \( R_2 \), of two characters of study in PSS over two occasions. The estimators, \( e_1 \) and \( e_2 \), were ratio-type and product-type estimators proposed along the line of Rao (1957) and Rao and Pereira (1968). The estimators proposed by Rao (1957) and Rao and Pereira (1968) were based on a complete matching of sample units on the second occasion. But, the estimators, \( e_1 \) and \( e_2 \), proposed by Onyeaka (2002) were based on a partial matching or replacement of sample units on the second occasion. This was in line with the work of Tripathi and Sinha (1976) who also based their estimator on partial replacement of sample units on the second occasion. However, the estimator proposed by Tripathi and Sinha (1976), unlike those proposed by Onyeaka (2002), was a ratio of two regression-type estimators.

Tripathi and Sinha (1976) actually worked on estimation of population ratio in a unistage sampling scheme over successive occasions. Their matched estimator of the second occasion population ratio, \( R_2 = \frac{Y_2}{X_2} \), was a ratio of a regression-type estimator of \( Y_2 \) to another regression-type estimator of \( X_2 \) where \( y \) and \( x \) are two characters of study. Okafor (1985), studying the use of multistage sampling over two occasions, proposed a wider variety of estimators of \( R_2 \) along the line of Tripathi and Sinha (1976). His estimators based
on the matched sample were different combinations of ratio-type, product-type and
difference-type double sampling estimators of \( R_2 \).

In the present study, we shall propose two estimators of \( R_2 \) along the line of Tripathi and
Sinha (1976) and Okafor (1985). The first estimator, \( d_1 \), is a ratio-cum-product type
estimator, while the second estimator, \( d_2 \), is a product-cum-ratio type estimator for the
matched part of the sample. Both estimators are linear functions of estimators based on the
matched and unmatched sample units. Properties of the proposed estimators, including the
estimators that provide the best linear combinations of the matched and unmatched estimators,
shall be obtained for repeated samples of fixed sizes. The performance of both estimators,
in terms of increased efficiency in estimating the population ratio, \( R_2 \), shall be
considered.

The Proposed Estimators

Consider the following sampling design for poststratified sampling over two occasions
proposed by Onyeka (2001).

A random sample of size \( n \) is drawn from a population of \( N \) units using simple random
sampling without replacement (SRSWOR) method on the first occasion. The sampled units
are allocated to their respective strata where \( n_{1h} \) is the number of units that fall into the \( h^{th} \)
stratum such that \( \sum_h n_{1h} = n, \ ( h = 1, 2, \ldots, L ) \). It is assumed that \( n \) is large enough such
that \( \text{Prob} ( n_{1h} = 0 ) = 0 \) for all \( h \). On the second occasion, \( m_h = \lambda n_{1h} \) units of the first
occasion sample are retained in the \( h^{th} \) stratum, \( \sum_h m_h = m = \lambda n, \ ( 0 < \lambda < 1 ) \). The
remaining \( u_{1h} = n_{1h} - m_h = n_{1h} - \lambda n_{1h} = \mu n_{1h} \) units are discarded, \( \sum_h u_{1h} = u = \mu n \), and \( \mu + \lambda = 1 \).

Then, the matched sample of size \( m \) is supplemented with a fresh (unmatched) sample of \( u \)
units drawn independently from the entire population, again using SRSWOR method. The \( u \)
sampled units are allocated to their respective strata where \( u_{2h} \) is the number of units that fall
into the \( h^{th} \) stratum such that \( \sum_h u_{2h} = u \left( = \sum_h u_{1h} \right) \). Again, it is assumed that \( u \) is large
enough such that \( \text{Prob} ( u_{2h} = 0 ) = 0 \) for all \( h \).

Let \( y_{ijh} \) and \( x_{ijh} \) denote observations on the \( i^{th} \) unit of the two characters of study in the \( h^{th} \)
stratum on the \( j^{th} \) occasion, \( i = 1, 2, \ldots, N; h = 1, 2, \ldots, L \) and \( j = 1, 2 \). The variate, \( x \), in
some cases can be the stratification variate. But generally, the variates \( y \) and \( x \) are to be
taken as any two characters of study. Let \( R_2 = \bar{Y}_2 / \bar{X}_2 \) and \( R_1 = \bar{Y}_1 / \bar{X}_1 \) respectively
denote the second and first occasion population ratios of the two characters of study. We
propose the following two estimators of \( R_2 \) in PSS over two occasions

\[
d_1 = \theta_1 \left( \frac{\bar{y}_{2m}}{\bar{y}_{1m}} \frac{\bar{x}_{2m}}{\bar{x}_{1m}} \right) + (1 - \theta_1)(\bar{y}_{2u} / \bar{x}_{2u})
\]
\[
\theta_1 \left( \frac{\bar{y}_{2m} \cdot \bar{y}_{1m} - x_{1h}}{\bar{y}_{1m} \cdot \bar{x}_{2m} - x_{1h}} \right) + (1 - \theta_1) \left( \frac{\bar{y}_{2u} - x_{1u}}{\bar{x}_{2u} - x_{1u}} \right) = \theta_1 d_{1m} + (1 - \theta_1) y_{2u} \tag{2.1}
\]

and
\[
\theta_2 \left( \frac{\bar{y}_{2m} \cdot \bar{y}_{1m} - x_{1h}}{\bar{y}_{1m} \cdot \bar{x}_{2m} - x_{1h}} \right) + (1 - \theta_2) \left( \frac{\bar{y}_{2u} - x_{1u}}{\bar{x}_{2u} - x_{1u}} \right) = \theta_2 d_{2m} + (1 - \theta_2) y_{2u} \tag{2.2}
\]

where
\[
\begin{align*}
\bar{y}_{2m} &= \sum_h W_h \bar{y}_{2h} \quad \bar{y}_{1m} = \sum_h W_h \bar{y}_{1h} \quad \bar{y}_{2u} = \sum_h W_h y^*_{2h} \\
\bar{x}_{2m} &= \sum_h W_h \bar{x}_{2h} \quad \bar{x}_{1m} = \sum_h W_h \bar{x}_{1h} \quad \bar{x}_{2u} = \sum_h W_h x^*_{2h}
\end{align*}
\]

\(\bar{y}_{1h}, \bar{x}_{1h}, \bar{y}_{2h}, \bar{x}_{2h}\) are sample means based on the entire first occasion sample of size \(n_{1h}\)
\(\bar{y}_{1h}^*, \bar{y}_{2h}^*, \bar{x}_{1h}^*, \bar{x}_{2h}^*\) are sample means based on the matched sample of size \(m_h\)
\(\bar{y}_{2h}^*, \bar{x}_{2h}^*\) are sample means based on the second occasion unmatched sample of size \(u_{2h}\)

and \(\theta_1\) and \(\theta_2\) are constant (weighting) fractions of the matched and unmatched parts of the estimators \(d_1\) and \(d_2\).

It should be noted that the estimators, \(d_1\) and \(d_2\), do not assume knowledge of the first occasion population ratio, \(R_1\). This is unlike the estimators proposed by Rao (1957) and Rao and Pereira (1968), and one of the estimators (\(e_1\)) proposed by Onyeka (2002). Again, the estimator, \(d_1\), is a ratio-cum-product type estimator since its matched estimator, \(d_{1m}\), is a ratio of a ratio-type double sampling estimator of \(Y\) to a product-type double sampling estimator of \(X\). Similarly, the estimator, \(d_2\), is a product-cum-ratio type estimator since its matched estimator, \(d_{2m}\), is a ratio of product-type double sampling estimator of \(Y\) to a ratio-type double sampling estimator of \(X\). The two estimators, \(d_1\) and \(d_2\), as already noted above are proposed along the line of Tripathi and Sinha (1976) and Okafor (1985).

Properties of the Proposed Estimators

Let the population variances of the study variates \(y\) and \(x\) in the \(h^{th}\) stratum be respectively denoted by \(S^2_{yh}\) and \(S^2_{xh}\) on both occasions. Also, let the covariance of \(y\) and \(x\) on both first and second occasions in the \(h^{th}\) stratum be denoted by \(S_{yhxh}\). Then, the properties of the proposed estimator, \(d_1\), are stated in Theorem 1, while those of the proposed estimator, \(d_2\), are stated in Theorem 2.
Theorem 1

The proposed ratio-cum-product estimator, $d_1 = 1 - (\bar{y}_{in} \cdot \bar{x}_{in} - x_{in}) \cdot (1 - \theta_1)(\bar{y}_{2n} / \bar{x}_{2n})$, is biased for the second occasion population ratio, $R_2$, in poststratified sampling over two occasions. For repeated samples of fixed sizes $n, m$ and $u$, the optimum value of the weighting fraction, $\theta_1$, and the associated mean square error of $d_1$ are respectively given by

$$\theta_1 = \frac{\lambda \sum_{h} W_h \sigma_{sh}}{\sum_{h} W_h \sigma_{sh} + \mu_1 R_2 \sum_{h} W_h \sigma_{1h} + 2 \mu_1 R_1 \sum_{h} W_h \sigma_{sh}} \quad (3.1)$$

and

$$\text{MSE}(d_1) = \frac{\sum_{h} W_h \sigma_{2h} + \mu_1 R_2 \sum_{h} W_h \sigma_{1h} - 2 \mu_1 R_1 \sum_{h} W_h \sigma_{sh}}{\sum_{h} W_h \sigma_{2h} + \mu_1 R_2 \sum_{h} W_h \sigma_{1h} - 2 \mu_1 R_1 \sum_{h} W_h \sigma_{sh}} \frac{\sum_{h} W_h \sigma_{2h}}{n X_2} \quad (3.2)$$

where

$$\begin{align*}
\sigma_{2h} &= S_{2h}^2 + \mu_1 R_2 S_{1h}^2 - 2 R_2 S_{2sh} \\
\sigma_{1h} &= S_{1h}^2 + \mu_2 R_2 S_{2sh}^2 + 2 R_2 S_{2sh}
\end{align*}$$

$$\sigma_{2h} = S_{2h}^2 - R_2 R_1 S_{2sh}^2 - (R_2 - R_1) S_{2sh}$$

and

$$R_2 = \frac{\bar{Y}_2}{\bar{X}_2}, \quad R_1 = \frac{\bar{Y}_1}{\bar{X}_1}, \quad R_y = \frac{\bar{Y}_2}{\bar{Y}_1}$$

Proof

From equation (2.1), the mean square error of the proposed ratio-cum-product estimator, $d_1$, can be written as:

$$\text{MSE}(d_1) = \theta_1^2 \text{MSE}(d_{1m}) + (1 - \theta_1)^2 \text{MSE}(\gamma_{2u}) + 2 \theta_1 (1 - \theta_1) C(d_{1m}, \gamma_{2u}) \quad (3.4)$$

where $\gamma_{2u} = \bar{y}_{2u} / \bar{x}_{2u}$ and $d_{1m} = \frac{\bar{y}_{2m} \cdot \bar{y}_{im} \cdot \bar{x}_{im}}{\bar{y}_{im} \cdot \bar{x}_{2m} \cdot \bar{x}_{im}} \quad (3.5)$

Using the least square method to minimize equation (3.4) with respect to $\theta_1$, the optimum value of the weighting fraction, $\theta_1$, and the associated mean square error of $d_1$ are respectively obtained as:

$$\theta_1 = \frac{\text{MSE}(\gamma_{2u})}{\text{MSE}(\gamma_{2u}) + \text{MSE}(d_{1m})} \quad (3.6)$$

and

$$\text{MSE}(d_1) = \frac{\text{MSE}(\gamma_{2u}) \text{MSE}(d_{1m})}{\text{MSE}(\gamma_{2u}) + \text{MSE}(d_{1m})} \quad (3.7)$$
It should be noted that the covariance of \( d_{1m} \) and \( \gamma_{2m} \) is zero since the two estimators are based on entirely independent matched and unmatched sample units. Onyeka (2002) obtained the mean square error of the unmatched estimator, \( \gamma_{2u} \), as:

\[
\text{MSE}(\gamma_{2u}) = (\mu_m X'_i) \sum_h W_h \sigma_{sh} \tag{3.8}
\]

where \( \mu_{2h} \) is as given in equation (3.3)

If we use the Taylor's series to expand the expression for \( d_{1m} \) in equation (3.5) up to terms of order \( n^{-1} \) in expected value, we have:

\[
(d_{1m} - R) = R_z (\delta^{2\bar{y}_{2m}} + \delta^{2\bar{y}_{1m}} + \delta^{2\bar{x}_{1m}} + \delta^{2\bar{x}_{2m}} + \delta^{2\bar{x}_{im}})
\]

\[
+ 2\delta^{2\bar{y}_{2m}} \delta y_{1m} + 2\delta^{2\bar{y}_{2m}} \delta x_{1m} - 2\delta^{2\bar{y}_{1m}} \delta y_{2m} - 2\delta^{2\bar{y}_{1m}} \delta x_{2m}
\]

\[
+ 2\delta^{2\bar{y}_{2m}} \delta x_{1m} - 2\delta^{2\bar{x}_{1m}} \delta y_{2m} - 2\delta^{2\bar{x}_{1m}} \delta x_{2m} - 2\delta^{2\bar{x}_{1m}} \delta y_{im} - 2\delta^{2\bar{x}_{1m}} \delta x_{im}
\]

\[
- 2\delta^{2\bar{x}_{1m}} \delta x_{im} + 2\delta^{2\bar{x}_{1m}} \delta x_{im} + 2\delta^{2\bar{x}_{1m}} \delta x_{im}
\]

where \( \delta^{2\bar{y}_{2m}} = \frac{\bar{y}_{2m} - \bar{y}_2}{\bar{y}_2} \), \( \delta^{2\bar{x}_{1m}} = \frac{\bar{x}_{1m} - \bar{x}_2}{\bar{X}_2} \), etc. Taking the conditional and unconditional expectations of equation (3.9) in sequence gives the unconditional mean square error of \( d_{1m} \) as:

\[
\text{MSE}(d_{1m}) = R_z^2 \left[ V(\bar{y}_{2m}) + (\bar{X}_2) V(\bar{y}_{1m}) + (\bar{X}_2) V(\bar{x}_{im})
\]

\[
+ 2(\bar{Y}_2, \bar{Y}_1) C(\bar{y}_{2m}, \bar{y}_{1m}) + 2(\bar{Y}_2, \bar{X}_1) C(\bar{y}_{2m}, \bar{x}_{1m}) - 2(\bar{Y}_2, \bar{Y}_1) C(\bar{x}_{2m}, \bar{x}_{1m}) - 2(\bar{Y}_2, \bar{Y}_1) C(\bar{x}_{2m}, \bar{x}_{1m})
\]

\[
- 2(\bar{Y}_2, \bar{X}_2) C(\bar{y}_{2m}, \bar{x}_{2m}) - 2(\bar{Y}_2, \bar{X}_2) C(\bar{y}_{2m}, \bar{x}_{2m}) - 2(\bar{Y}_2, \bar{X}_2) C(\bar{y}_{2m}, \bar{x}_{2m}) - 2(\bar{Y}_2, \bar{X}_2) C(\bar{y}_{2m}, \bar{x}_{2m})
\]

\[
- 2(\bar{Y}_2, \bar{X}_2) C(\bar{y}_{2m}, \bar{x}_{2m}) + 2(\bar{Y}_2, \bar{X}_2) C(\bar{y}_{2m}, \bar{x}_{2m}) + 2(\bar{Y}_2, \bar{X}_2) C(\bar{y}_{2m}, \bar{x}_{2m})
\]

where \( V \) and \( C \) respectively represent unconditional variance and covariance for repeated samples of fixed sizes \( n \) and \( m \). Following Ige (1984) and Onyeka (2007), we have:

\[
V(\bar{y}_{2m}) = (\lambda n) \sum_h W_h S_{sh}^2, \ V(\bar{y}_{2m}) = (\lambda n) \sum_h W_h S_{sh}^2, \ C(\bar{y}_{2m}, \bar{x}_{2m}) = (\lambda n) \sum_h W_h S_{xyh}, \ etc.
\]

And, on making the necessary substitutions in equation (3.10), we obtain the mean square error of the matched estimator, \( d_{1m} \), as:

\[
\text{MSE}(d_{1m}) = (\lambda n \bar{X}_2)^{-1} \left[ \sum_h W_h \sigma_{2h} + \mu R_y^2 \sum_h W_h \sigma_{sh} - 2 \mu R_y \sum_h W_h \sigma_{sh} \right] \tag{3.11}
\]

where \( \mu_2h, \mu_1h \) and \( \mu_3h \) are as given in equation (3.3). The optimum weighting fraction, \( \mu_{01} \),
and the associated mean square error of the proposed estimator, \( d_1 \), as given in the theorem, are obtained by using equations (3.8) and (3.11) to make the necessary substitutions in equations (3.6) and (3.7). This completes the proof.

**Theorem 2**

The proposed product-cum-ratio estimator, \( d_1 = \left[ y_{im} \cdot \bar{y}_{1m} \cdot \bar{x}_{im} \right] + (1 - 0.2)(\bar{y}_{2m} / \bar{x}_{2m}) \), is biased for the second occasion population ratio, \( R_{2i} \), in poststratified sampling over two occasions. For repeated samples of fixed sizes \( n, m \) and \( u \), the optimum value of the weighting fraction, \( \lambda_{2h} \), and the associated mean square error of \( d_2 \) are respectively given by:

\[
\lambda_{2h} = \frac{\sum h W_h \sigma_{2h}}{\sum h W_h \sigma_{2h} + \mu R \sum h W_h \sigma_{1h} + 2 \mu R \sum h W_h \sigma_{3h}} \tag{3.12}
\]

and

\[
\text{MSE}(d_2) = \frac{\sum h W_h \sigma_{2h} + \mu R \sum h W_h \sigma_{1h} + 2 \mu R \sum h W_h \sigma_{3h} + \sum h W_h \sigma_{2h}}{\sum h W_h \sigma_{2h} + \mu R \sum h W_h \sigma_{1h} + 2 \mu R \sum h W_h \sigma_{3h} + n \bar{X}_2^2} \tag{3.13}
\]

where \( \Omega_{2h} \), \( \Theta_{1h} \) and \( \Theta_{3h} \) are as already defined in equation (3.3).

**Theorem 2** can be proved by following similar steps as in the proof of Theorem 1.

**Optimum Replacement Fraction**

One of the major issues often considered in successive sampling is the replacement policy or the manner in which the sample should be changed on subsequent occasions. A complete matching or replacement of sample units is often recommended when estimating changes. An entirely new sample or no matching sample is favoured when estimating average or sum of population parameters over successive occasions. A partial replacement or matching of sample units is preferred when the interest is on the current occasion estimates. But, with a partial matching of sample units, like in the present study, it is often needful to obtain an expression for the optimum matching fraction. Such a value would definitely minimize the variance or mean square error of the current occasion estimates. Theorem 3 gives the optimum matching or replacement fractions for the two estimators, \( d_1 \) and \( d_2 \), proposed in this study.

**Theorem 3**

The optimum replacement or matching fraction of the proposed ratio-cum-product type estimator, \( d_1 \), is given by:

\[
\lambda_{o1} = \frac{1 + \Lambda_1 - \sqrt{1 + \Lambda_1}}{\Lambda_1} \tag{4.1}
\]
\[
\sum_{h} W_h S_{yh} > R \sum_{h} W_h S_{ysh} + (R - 1) \sum_{h} W_h S_{xsh} \tag{5.1}
\]

while the estimator \( d_2 \) would perform better than \( d_1 \) if

\[
\sum_{h} W_h S_{yh} < R \sum_{h} W_h S_{ysh} + (R - 1) \sum_{h} W_h S_{xsh} \tag{5.2}
\]

Theorem 4 can be proved by considering the difference of the optimum mean square errors of \( d_1 \) and \( d_2 \) as respectively given in equations (4.2) and (4.5).

In summary, it is clear from Theorem 4 that none of the two proposed estimators always performs better than the other, in terms of having a smaller mean square error. The efficiency of one estimator over the other can only be established under certain conditions stated in Theorem 4. However, an added advantage of both estimators over many other ratio estimators is that the two proposed estimators do not require knowledge of the first occasion population ratio, \( R \). This, invariably, eliminates a great deal of material and financial resources that would have been devoted to the procurement of information on \( R \), as it is often the case with estimators that require knowledge of the first occasion population ratio, \( R \).

References

Okafor, F. C. 1985. Use of multistage sampling over two occasions. A Ph.D. thesis submitted to the Faculty of Science, University of Ibadan, Nigeria.


and the associated optimum mean square error of \( d_1 \) is

\[
\text{MSE}_n(d_1) = (2n \bar{X}_2) \left( 1 + \sqrt{1 + A_1} \right) \sum_h W_h \sigma_{zh} \tag{4.2}
\]

where

\[
A_1 = \frac{R^2 \sum_h W_h \sigma_{1h} + 2R \sum_h W_h \sigma_{2h}}{\sum_h W_h \sigma_{zh}} \tag{4.3}
\]

Furthermore, the optimum replacement or matching fraction of the proposed product-cum-ratio type estimator, \( d_2 \), is given by:

\[
\lambda_{02} = \frac{1 + A_2 - \sqrt{1 + A_2}}{A_2} \tag{4.4}
\]

and the associated optimum mean square error of \( d_2 \) is

\[
\text{MSE}_n(d_2) = (2n \bar{X}_2) \left( 1 + \sqrt{1 + A_2} \right) \sum_h W_h \sigma_{zh} \tag{4.5}
\]

where

\[
A_2 = \frac{R^2 \sum_h W_h \sigma_{1h} + 2R \sum_h W_h \sigma_{2h}}{\sum_h W_h \sigma_{zh}} \tag{4.6}
\]

**Proof**

Using the least square method to minimize equation (3.2) with respect to \( \lambda \) gives the optimum value of \( \lambda \) for the ratio-cum-product type estimator, \( d_1 \), as:

\[
\lambda_{01} = \frac{1}{1 + \sqrt{1 + A_1}} \tag{4.1}
\]

where \( A_1 \) is as given in equation (4.3). The required optimum matching fraction (equation (4.1)), is then obtained as \( \lambda_{01} = 1 - \lambda_{01} \). Also, the associated optimum mean square error (equation (4.2)) of the proposed ratio-cum-product type estimator; \( d_1 \), is obtained by substituting \( \lambda \) with \( \lambda_{01} \) in equation (3.2). Similarly, using the least square method to minimize equation (3.13) with respect to \( \lambda \) leads to equations (4.4) and (4.5) as stated in the theorem. And this completes the proof.

**Comparison of Proposed Estimators**

The ratio-cum-product type estimator, \( d_1 \), and the product-cum-ratio type estimator, \( d_2 \), proposed in the present study are both biased for the second occasion population ratio, \( R_2 \), in poststratified sampling over two occasions. The following theorem compares the performance and efficiency of both estimators in terms of the estimator with the smaller mean square error.

**Theorem 4**

The proposed ratio-cum-product type estimator, \( d_1 \), would perform better than the proposed product-cum-ratio type estimator, \( d_2 \), in terms of having a smaller mean square error, if