Estimating a finite population mean under random non-response in two stage cluster sampling with replacement

Nelson Kiprono Bii$^{1,a}$, Christopher Ouma Onyango$^{2}$ and John Odhiambo$^{1}$

$^1$Institute of Mathematical Sciences, Strathmore University, P.O Box 59857-00200, Nairobi, Kenya
$^2$Department of Statistics, Kenyatta University, P.O Box 43844-00200, Nairobi, Kenya
E-mail: $^a$nkiprono@strathmore.edu

Abstract. Non-response is a regular occurrence in sample surveys. Developing estimators when non-response exists may result in large biases during estimation of population parameters. In this paper, a finite population mean is estimated when non-response exists randomly under two stage cluster sampling with replacement. It is assumed that non-response arises in the survey variable in the second stage of cluster sampling. Weighting method of compensating for non-response is applied. Asymptotic properties of the proposed estimator of the population mean are derived. Under mild assumptions, the estimator is shown to be asymptotically consistent.

1. Introduction

In survey sampling, non-response is one source of errors in data analysis. Non-response introduces bias into the estimation of population characteristics. It also causes samples fail to follow the distributions determined by the original sampling design. This paper seeks to reduce the non-response bias in the estimation of a finite population mean in two stage cluster sampling. Use of regression models is recognized as one of the procedures for reducing bias due to non-response using auxiliary information. In practice, information on the variables of interest is not available for non-respondents but information on auxiliary variables may be available for non-respondents. It is therefore desirable to model the response behavior and incorporate the auxiliary data into the estimation so that the bias arising from non-response can be reduced. If the auxiliary variables are correlated with the response behavior, then the regression estimators would be more precise in estimation of population parameters, given the auxiliary information is known. Many authors have developed estimators of population mean where non-response exists in the study and auxiliary variables. But there exist cases that do not exhibit non-response in the auxiliary variables, such as; number of people in a family, duration one takes to go through education. Imputation techniques have been used to account for non-response in the study variable. For instance, [7] applied compromised method of imputation to estimate a finite population mean under two stage cluster sampling, this method however produced a large bias. In this study, the Nadaraya-Watson regression technique is applied in deriving the estimator for the finite population mean. Kernel weights are used to compensate for non-response.
2. Reweighting method

Non-response causes loss of observations and therefore reweighting means that the weights are increased for all or almost all of the elements that fail to respond in a survey. The population mean, \( \bar{Y} \), is estimated by selecting a sample of size \( n \) at random with replacement. If responding units of item \( y \) are independent so that the probability of unit \( j \) responding in cluster \( i \) is \( p_{ij}(i = 1, 2, \cdots, n; j = 1, 2, \cdots, m) \) then an imputed estimator, \( \hat{y}_i \), for \( \bar{Y} \), is given by

\[
\hat{y}_i = \frac{1}{\sum_{i,j \in s} w_{i,j}} \left[ \sum_{i,j \in s_r} w_{i,j}y_{i,j} + \sum_{i,j \in s_m} w_{i,j}y^*_{i,j} \right],
\]

where \( w_{i,j} = \frac{1}{\pi_{ij}} \) gives the survey weight tied to unit \( j \) in cluster \( i \) and \( \pi_{ij} = p[i, j \in s] \) is its second order probability of inclusion, \( s_r \), is the set of \( r \) units responding to item \( y \) so that \( r + m = n \) and \( y^*_{i,j} \) is the imputed value generated so that the missing value \( y_{i,j} \) is compensated for [6].

3. The Proposed Estimator of Finite Population Mean

Consider a finite population of size \( M \) consisting of \( N_i \) elements in the \( i^{th} \) cluster. A sample of \( n \) clusters is selected so that \( n_1 \) units respond and \( n_2 \) units fail to respond. Let \( y_{i,j} \) denote the value of the survey variable \( y \) for unit \( j \) in cluster \( i \), for \( i = 1, 2, \cdots, N, j = 1, 2, \cdots, N_i \) and let population mean be given by

\[
\bar{Y} = \frac{1}{MN_i} \sum_{i=1}^{N} \sum_{j=1}^{M_i} Y_{ij}.
\]

Let an estimator of the finite population mean be defined by \( \hat{Y} \) as follows

\[
\hat{Y} = \frac{1}{M} \left\{ \frac{1}{n_1} \sum_{i \in s} \sum_{j \in s} Y_{ij} \delta_{ij} + \frac{1}{n_2} \sum_{i \in s} \sum_{j \notin s} \left( 1 - \frac{1}{\pi_{ij}} \right) \hat{Y}_{ij} \delta_{ij} \right\}.
\]

where \( \delta_{ij} \) is an indicator variable defined by

\[
\delta_{ij} = \begin{cases} 1, & \text{if } j^{th} \text{ unit in the } i^{th} \text{ cluster responds} \\ 0, & \text{elsewhere} \end{cases}
\]

and \( n_1 \) and \( n_2 \) are the number of units that respond and those that fail to respond respectively. \( \pi_{ij} \) is the probability of selecting the \( j^{th} \) unit from the \( i^{th} \) cluster into the sample. Let \( w(x_{ij}) = \frac{1}{\pi_{ij}} \) to be the inverse of the second order inclusion probabilities and \( x_{ij} \) be the \( j^{th} \) auxiliary random variable from the \( i^{th} \) cluster. It follows that equation (3.2) becomes

\[
E(\hat{Y}) = \frac{1}{M} \left\{ \frac{1}{n_1} \sum_{i \in s} \sum_{j \in s} w(x_{ij}) \delta_{ij} Y_{ij} + \frac{1}{n_2} \sum_{i \in s} \sum_{j \notin s} \left( 1 - w(x_{ij}) \right) \hat{Y}_{ij} \delta_{ij} \right\}.
\]

Suppose \( \delta_{ij} \) is known to be Bernoulli random variables with probability of success \( \delta^*_j \), then, \( E(\delta_{ij}) = p_r(\delta_{ij} = 1) = \delta^*_j \) and \( \text{Var}(\delta) = \delta^*_j(1 - \delta^*_j) \), [1]. Thus, the expected value of the estimator of population mean is given by
\[
E(\hat{Y}) = \frac{1}{M} \left\{ \frac{1}{n_1} \sum_{i \in s} \sum_{j \in s} E(w(x_{ij})Y_{ij}) \delta_{ij} + \frac{1}{n_2} \sum_{i \in s} \sum_{j \notin s} E((1 - w(x_{ij}))\hat{Y}_{ij}) \delta^{*}_{ij} \right\}.
\] (3.5)

Assuming non-response in the second stage of sampling, the problem is therefore to estimate the values of \(\hat{Y}_{ij}\). To do this, a linear regression model applied by [5] and [4] given below is used:

\[
\hat{Y}_{ij} = m(\hat{x}_{ij}) + \hat{e}_{ij}.
\] (3.6)

where \(m(.)\) is as smooth function of the auxiliary variables and \(\hat{e}_{ij}\) is the residual term with mean zero and variance which is strictly positive. Substituting equation (3.6) in equation (3.5), the following result is obtained:

\[
E(\hat{Y}) = 1 \left\{ \frac{M}{n_1} \sum_{i \in s} \sum_{j \in s} \frac{E((m(\hat{x}_{ij}) + \hat{e}_{ij})w(x_{ij}))}{\delta_{ij}} \right\}.
\] (3.7)

Assuming \(n_1 = n_2 = n\), and simplifying equation (3.7), we obtain the following

\[
E(\hat{Y}) = \frac{1}{Mn} \left\{ \sum_{i \in s} \sum_{j \in s} E((m(\hat{x}_{ij}) + \hat{e}_{ij})w(x_{ij})) \delta_{ij} + \sum_{i \in s} \sum_{j \notin s} E((1 - w(x_{ij}))(m(\hat{x}_{ij}) + \hat{e}_{ij})) \delta^{*}_{ij} \right\}.
\] (3.8)

A detailed work done by [4] proved that \(E(\hat{e}_{ij}) = 0\). Therefore equation (3.8) reduces to

\[
E(\hat{Y}) = \frac{1}{Mn} \left\{ \sum_{i \in s} \sum_{j \in s} E(m(\hat{x}_{ij}))E(w(x_{ij})) \delta_{ij} + \sum_{i \in s} \sum_{j \notin s} E((1 - w(x_{ij}))(m(\hat{x}_{ij})) \delta^{*}_{ij} \right\}.
\] (3.9)

The second term in equation (3.9) is simplified as follows:

\[
\frac{1}{Mn} \left\{ \sum_{i \in s} \sum_{j \notin s} E((1 - w(x_{ij}))(m(\hat{x}_{ij}) + \hat{e}_{ij}) \delta^{*}_{ij} \right\} = \frac{1}{Mn} \left\{ \sum_{i \in s} \sum_{j \notin s} E((1 - w(x_{ij}))m(\hat{x}_{ij})) \delta_{ij} \right\}
\]

\[+ \frac{1}{Mn} \left\{ \sum_{i \in s} \sum_{j \notin s} E((1 - w(x_{ij}))) \delta_{ij} \hat{e}_{ij} \right\}.
\] (3.10)

However, \(E(m(x_{ij})) = m(\hat{x}_{ij}) = m(x_{ij})\) [2]. Thus we get the following:
\[
\frac{1}{M_n} \left\{ \sum_{i \in s} \sum_{j \notin s} E(1 - w(x_{ij}))(m(x_{ij}) + e_{ij})\delta_{ij} \right\} = \frac{1}{M_n} \left\{ \sum_{i = n+1}^{N} \sum_{j = m+1}^{M} \delta_{ij}m(x_{ij}) - w(x_{ij})\delta_{ij}m(x_{ij}) \right\} + \frac{1}{M_n} \left\{ \sum_{i = n+1}^{N} \sum_{j = m+1}^{M} E(e_{ij}\delta_{ij}) - E(w(x_{ij})(e_{ij}\delta_{ij})) \right\}.
\]

Equation (3.11) then can be simplified as follows:

\[
\frac{1}{M_n} \left\{ \sum_{i \in s} \sum_{j \notin s} E(1 - w(x_{ij}))(m(x_{ij}) + e_{ij})\delta_{ij} \right\} = \frac{1}{M_n} \left\{ (M - (m + 1))(N - (n + 1)) \right. \\
\left. - (n + 1)[\delta_{ij}m(x_{ij}) - w(x_{ij})\delta_{ij}m(x_{ij})] + (M - (m + 1))(N - (n + 1)) \right\} \\
\left\{ \delta_{ij}E(e_{ij} - E(e_{ij}\delta_{ij}w(x_{ij})) \right\}.
\]

According to [4], \(E(e_{ij}) = 0\). Thus equation (3.12) can be reduced to

\[
\frac{1}{M_n} \left\{ \sum_{i \in s} \sum_{j \notin s} E(1 - w(x_{ij}))(m(x_{ij}) + e_{ij})\delta_{ij} \right\} = \frac{(M - (m + 1))(N - (n + 1))}{M_n} \times \Delta_1.
\]

where \(\Delta_1 = \left\{ \delta_{ij}m(x_{ij})(1 - w(x_{ij})) \right\} \). Recall that \(w(x_{ij}) = \frac{1}{\pi_{ij}}\) so that equation (3.13) may be re-written as follows:

\[
\frac{1}{M_n} \left\{ \sum_{i \in s} \sum_{j \notin s} E(1 - w(x_{ij}))(m(x_{ij}) + e_{ij})\delta_{ij} \right\} = \frac{(M - (m + 1))(N - (n + 1))}{M_n} \times \Delta_2.
\]

where \(\Delta_2 = \left\{ \delta_{ij}m(x_{ij}) \left( \frac{\pi_{ij} - 1}{\pi_{ij}} \right) \right\} \). Assume the sample sizes are large i.e. \(n \to N\) and \(m \to M\), then equation (3.14) can be simplified to

\[
\frac{1}{M_n} \left\{ \sum_{i \in s} \sum_{j \notin s} E(1 - w(x_{ij}))(m(x_{ij}) + e_{ij})\delta_{ij} \right\} = \frac{1}{M_n} \left\{ \delta_{ij}m(x_{ij}) \left( \frac{\pi_{ij} - 1}{\pi_{ij}} \right) \right\}.
\]

Combining equation (3.15) and the first term in equation (3.9), the following result is obtained:

\[
E(\hat{Y}) = \frac{1}{M_n} \left\{ \sum_{i \in s} \sum_{j \notin s} E(m(x_{ij})) \left( \frac{\delta_{ij}}{\pi_{ij}} \right) + \sum_{i \in s} \sum_{j \notin s} \delta_{ij} \left( \frac{\pi_{ij} - 1}{\pi_{ij}} \right) \right\}.
\]
Since the first term represents the response units, their values are all known. The problem is to estimate the non-response units in the second term. Let the indicator variable \( \delta_{ij} = 1 \), the problem now reduces to that of estimating the function \( m(x_{ij}) \), which is a function of the auxiliary variables, \( x_{ij} \).

Hence the expected value of the estimator of the finite population mean under non-response is given as:

\[
E(\hat{Y}) = \frac{1}{Mn} \left\{ \sum_{i \in s} \sum_{j \in s} Y_{ij} + \sum_{i \in s} \sum_{j \notin s} \delta_{ij} (m(x_{ij})) \left( \frac{\pi_{ij}}{\pi_{ij}} - 1 \right) \right\}, \tag{3.17}
\]

In order to derive the asymptotic properties of the expected value of the proposed estimator in equation (3.17), a review of Nadaraya-Watson estimator is given in the next section.

4. Review of Nadaraya-Watson Estimator

Given a random sample of bi-variate data \((x_i, y_i), \ldots, (x_n, y_n)\) having a joint pdf \(g(x, y)\) with the regression model given by \(Y_{ij} = m(x_{ij}) + e_{ij}\) as in equation (3.6), where \(m(.)\) is unknown.

It is assumed that the error term satisfies the following conditions:

\[
E(e_{ij}) = 0, \quad Var(e_{ij}) = \sigma^2_{ij}, \quad Cov(e_i, e_j) = 0, \quad \text{for } i \neq j. \tag{4.1}
\]

Furthermore, let \(K(.)\) denote a kernel density function which is twice continuously differentiable with:

\[
\begin{align*}
\int_{-\infty}^{\infty} k(w)dw &= 1 \\
\int_{-\infty}^{\infty} wk(w)dw &= 0 \\
\int_{-\infty}^{\infty} k^2(w)dw &< \infty \\
\int_{-\infty}^{\infty} w^2k(w)dw &= d_k \\
k(w) &= k(-w).
\end{align*}
\tag{4.2}
\]

In addition, let the smoothing weights be

\[
w(x_{ij}) = \frac{K \left( \frac{x - X_{ij}}{b} \right)}{\sum_{i \in s} \sum_{j \in s} K \left( \frac{x - X_{ij}}{b} \right)}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, m. \tag{4.3}
\]

where \(b\) is a smoothing parameter, normally referred to as the bandwidth such that \(\sum_{i \in s} \sum_{j \in s} = 1\).

Using equation (4.3), the Nadaraya-Watson estimator of \(m(x_{ij})\) is

\[
m(x_{ij}) = \sum_{i \in s} \sum_{j \in s} w(x_{ij}) Y_{ij} = \frac{\sum_{i \in s} \sum_{j \in s} K \left( \frac{x - X_{ij}}{b} \right) Y_{ij}}{\sum_{i \in s} \sum_{j \in s} K \left( \frac{x - X_{ij}}{b} \right)}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, m. \tag{4.4}
\]

Given the model \(Y_{ij} = m(x_{ij}) + e_{ij}\) and the conditions of the error term as explained in (4.1) above, the expression of the expected value of the survey variable \(Y_{ij}\) relative to the auxiliary variable \(X_{ij}\) can be given as a joint pdf of \(g(x_{ij}, y_{ij})\) as follows:

\[
m(x_{ij}) = E(Y_{ij}/X_{ij} = x_{ij}) = \int y \ g \ [y|x] \ dy. \tag{4.5}
\]
Equation (4.5) can therefore be given as

$$m(x_{ij}) = \frac{\int y g(x, y) dy}{\int g(x, y) dy}. \tag{4.6}$$

where $\int g(x, y) dy$ is the marginal density of $X_{ij}$. The numerator and the denominator of equation (4.6) can be estimated separately using kernel functions such as $g(x, y)$ is estimated by:

$$\hat{g}(x, y) = \frac{1}{mn} \sum_i \sum_j \left( \frac{1}{b} K\left( \frac{x - X_{ij}}{b} \right) \frac{1}{b} K\left( \frac{y - Y_{ij}}{b} \right) \right), \tag{4.7}$$

and

$$\int y \hat{g}(x, y) dy = \frac{1}{mn} \sum_i \sum_j \int \left( \frac{1}{b} K\left( \frac{x - X_{ij}}{b} \right) \frac{1}{b} K\left( \frac{y - Y_{ij}}{b} \right) \right) y dy. \tag{4.8}$$

Change of variables technique is used so that making $y$ the subject and differentiating it with respect to $w$ gives $dy = bw$ as outlined in equation (4.9)

$$w = \frac{y - Y_{ij}}{b},$$

$$y = wb + Y_{ij},$$

$$dy = bw. \tag{4.9}$$

so that

$$\int y \hat{g}(x, y) dy = \frac{1}{mn} \sum_i \sum_j \int \frac{1}{b} K\left( \frac{x - X_{ij}}{b} \right) \frac{1}{b} K\left( \frac{bw + Y_{ij}}{b} \right) (bw + Y_{ij}) K(bw) dy. \tag{4.10}$$

Equation (4.10) can be reduced to

$$\int y \hat{g}(x, y) dy = \frac{1}{mn} \sum_i \sum_j K\left( \frac{x - X_{ij}}{b} \right) \left[ \int w K(w) bw + \frac{1}{b} Y_{ij} \int K(w) bw \right]. \tag{4.11}$$

From the conditions specified in equation (4.2), equation (4.11) can be simplified to

$$\int y \hat{g}(x, y) dy = \frac{1}{mn} \sum_i \sum_j K\left( \frac{x - X_{ij}}{b} \right) Y_{ij}. \tag{4.12}$$

Following the same procedure, the denominator of equation (4.6) can also be simplified as

$$\int g(x, y) dy = \frac{1}{mn} \sum_i \sum_j \int \left( \frac{1}{b} K\left( \frac{x - X_{ij}}{b} \right) \frac{1}{b} K\left( \frac{y - Y_{ij}}{b} \right) \right) dy. \tag{4.13}$$

Re-writing equation (4.13) gives

$$\int g(x, y) dy = \frac{1}{mn} \sum_i \sum_j \int K\left( \frac{x - X_{ij}}{b} \right) \frac{1}{b} K\left( \frac{y - Y_{ij}}{b} \right) dy. \tag{4.14}$$
Using change of variable technique as in equation (4.9), equation (4.14) can be re-written as

\[
\int g(\hat{x}, y)dy = \frac{1}{mn\hat{b}} \sum_{i} \sum_{j} K \left( \frac{x - X_{ij}}{b} \right) \int \frac{1}{b} K(w)bdw. \tag{4.15}
\]

which yields

\[
\int g(\hat{x}, y)dy = \frac{1}{mn\hat{b}} \sum_{i} \sum_{j} K \left( \frac{x - X_{ij}}{b} \right).	ag{4.16}
\]

Since \( \int \frac{1}{b} K(w)bdw \) is a probability density function and therefore integrates to 1.

It follows from equation (4.13) and (4.16) that the estimator \( m(x_{ij}) \) is as given in equation (4.4). Thus the estimator of \( m(x_{ij}) \) is a linear smoother since it is a linear function of the observations, \( Y_{ij} \). Given a sample and a specified kernel function, the corresponding \( y \) estimate is obtained for a given auxiliary variable \( x_{ij} \) by using the estimator outlined in equation (4.4), such as

\[
y_{ij} = m_{NW}(\hat{x}_{ij}) = \sum_{i} \sum_{j} W_{ij}(x_{ij})Y_{ij}, \tag{4.17}
\]

where \( m_{NW}(\hat{x}_{ij}) \) is the Nadaraya-Watson estimator for estimating the unknown function \( m(.) \).

See [3, 8] for the details.

This provides a way of estimating for instance the non-response values of the survey variable \( Y_{ij} \), given the auxiliary values \( x_{ij} \), for a specified kernel function.

5. Asymptotic Bias of the Mean estimator, \( \hat{Y} \)

Equation (3.17) may be written as

\[
E(\hat{Y}) = \frac{1}{Mn} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m} Y_{ij} + \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} m_{NW}(\hat{y}_{ij}) \right\}. \tag{5.1}
\]

Re-writing equation (4.17) using the property of symmetry associated with Nadaraya-Watson estimator gives

\[
m_{NW}(\hat{x}_{ij}) = \frac{\sum_{i \in s} \sum_{j \in s} K \left( \frac{X_{ij} - x_{ij}}{b} \right) Y_{ij}}{\sum_{i \in s} \sum_{j \in s} K \left( \frac{X_{ij} - x_{ij}}{b} \right)}, \quad i = 1, 2, \ldots, n; \ j = 1, 2, \ldots, m. \tag{5.2}
\]

Equation (5.2) can be simplified to

\[
m_{NW}(\hat{x}_{ij}) = \frac{1}{g(\hat{x}_{ij})} \left[ \frac{1}{mn\hat{b}} \sum_{i} \sum_{j} K \left( \frac{X_{ij} - x_{ij}}{b} \right) Y_{ij} \right]. \tag{5.3}
\]

where \( g(\hat{x}_{ij}) \) is the estimated marginal density of auxiliary variables \( X_{ij} \). For a finite population mean, the expected value of the estimator is given in equation (5.1). The bias is given by

\[
Bias(\hat{Y}) = E(\hat{Y} - \bar{Y}). \tag{5.4}
\]
\[ \text{Bias} (\hat{Y}) = E \left\{ \frac{1}{MN} \left[ \sum_{i=1}^{n} \sum_{j=1}^{m} Y_{ij} + \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} m(x_{ij}) \right] - \frac{1}{MN} \left[ \sum_{i=1}^{n} \sum_{j=1}^{m} Y_{ij} + \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} m(x_{ij}) \right] \right\}. \] (5.5)

which can be reduced to

\[ \text{Bias} (\hat{Y}) = \frac{1}{MN} E \left\{ \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} m(x_{ij}) - \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} Y_{ij} \right\}. \] (5.6)

Re-write the regression model given by

\[ Y_{ij} = m(X_{ij}) + e_{ij} \] as

\[ Y_{ij} = m(x_{ij}) + [m(X_{ij}) - m(x_{ij})] + e_{ij}. \] (5.7)

so that from equation (5.3) the first term in equation (5.6) before taking expectation is given as

\[ \frac{1}{MN} \left\{ \frac{1}{mnb} \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} K(\frac{X_{ij} - x_{ij}}{b}) Y_{ij} \right\} \]

\[ = \frac{1}{MN} g(x_{ij}) \left\{ \frac{1}{mnb} \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} K(\frac{X_{ij} - x_{ij}}{b}) m(x_{ij}) \right\} + \frac{1}{mnb} \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} K(\frac{X_{ij} - x_{ij}}{b}) e_{ij} \}. \] (5.8)

Equation (5.9) then is simplified as below

\[ \frac{1}{MN} \left\{ \frac{1}{mnb} \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} K(\frac{X_{ij} - x_{ij}}{b}) Y_{ij} \right\} \]

\[ = \frac{1}{MN} \left\{ \frac{1}{mnb} g(x_{ij}) \right\} \left\{ \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} g(x_{ij}) m(x_{ij}) \right\} \]

\[ + \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} m_{1}(x_{ij}) + \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} m_{2}(x_{ij}) \}. \] (5.9)

where

\[ m_{1}(x_{ij}) = \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} K(\frac{X_{ij} - x_{ij}}{b}) [m(X_{ij}) - m(x_{ij})]. \] (5.10)
\[
m_2(x_{ij}) = \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} K \left( \frac{X_{ij} - x_{ij}}{b} \right) e_{ij}.
\] (5.11)

Based on the conditional expectation of equation (5.9) we get

\[
E \left[ \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} M(x_{ij})/x_{ij} \right] = \frac{1}{MN} E \left[ \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} \left[ m(x_{ij}) + \frac{m_1(x_{ij})}{g(x_{ij})} + \frac{m_2(x_{ij})}{g(x_{ij})} \right] \right].
\] (5.12)

To obtain the relationship between the conditional mean and the selected bandwidth, the following theorem due to [2] is applied.

**Theorem:** [2].

Let \( K(w) \) be a symmetric density function with \( \int wk(w)dw = 0 \) and \( \int w^2k(w)dw = k_2 \). Assume \( n \) and \( N \) increase together such that \( \frac{n}{N} \to \pi \) with \( 0 < \pi < 1 \). Besides, assume the sampled and non-sampled values of \( x \) are in the interval \([c, d]\) and are generated by densities \( d_n \) and \( d_{p-s} \) respectively. Both are bounded away from zero in the interval \([c, d]\) and assumed to have continuous second derivatives. For any variable \( Z, E(Z/U = u) = A(u) + O(B) \) and \( Var(Z/U = u) = O(C) \), then \( Z = A(u) + O_p(B + C^2) \). Applying this theorem, we have

\[
MSE(\widehat{X}/x_{ij}) = \frac{1}{(MN)^2} \left\{ \frac{(MN - mn)^2}{mnb g(x_{ij})} \int k(w^2)dw - \frac{(MN - mn)^2}{mnb g(x_{ij})} \int k(w^2)dw - \frac{g(x_{ij})^2}{g(x_{ij})} \right\} \left[ m''(x_{ij}) \right]
\]

\[
+ O(b^4) + \left[ \frac{(MN - mn)^2}{mnb} + \frac{1}{mnb} \right]^2.
\] (5.13)

This theorem has been stated without proof in the literature. To prove it, we partition it into bias and variance terms and separately prove them as follows:

From the conditions stated in (4.1) it follows that \( E(e_{ij}/X_{ij}) = 0 \). Therefore, \( E[m_2(x_{ij})] = 0 \). Thus, \( E[m_1(x_{ij})] \) can be obtained as follows:

\[
E \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} [m_1(x_{ij})] = \frac{1}{MN} \left\{ \frac{1}{mnb} \right\} E \left\{ \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} K \left( \frac{X_{ij} - x_{ij}}{b} \right) [m(X_{ij}) - m(x_{ij})] \right\}.
\] (5.14)

The substitution \( w = \frac{V - x_{ij}}{b} \) is used then \( V \) is differentiated with respect to \( w \) as indicated in equation (5.15)

\[
w = \frac{V - x_{ij}}{b}
\]

\[
V = x_{ij} + bw
\]

\[
dV = bw dw.
\] (5.15)
Equation (5.14) can then be simplified as follows

\[
E \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} [m_1(x_{ij})] = \frac{1}{MN} \left\{ \begin{array}{c}
MN - mn \\
mnb
\end{array} \right\} \int k(w) [m(x_{ij} + bw) - m(x_{ij})] \\
\times \int g(x_{ij} + bw) bw dw \}.
\] (5.16)

Using Taylor’s series expansion about the point \(x_{ij}\), the \(k^{th}\) order kernel can be derived as follows:

\[
g(x_{ij} + bw) = g(x_{ij}) + g'(x_{ij})bw + \frac{1}{2} g''(x_{ij})b^2 w^2 + \cdots + \frac{1}{k!} g^k(x_{ij})b^k w^k + O(b^2).
\] (5.17)

Similarly,

\[
m(x_{ij} + bw) = m(x_{ij}) + m'(x_{ij})bw + \frac{1}{2} m''(x_{ij}) b^2 w^2 + \cdots + \frac{1}{k!} m^k(x_{ij})b^k w^k + O(b^2).
\] (5.18)

Expanding up to the 3rd order kernels, equation (5.18) becomes

\[
[m(x_{ij} + bw) - m(x_{ij})] = m'(x_{ij})bw + \frac{1}{2} m''(x_{ij}) b^2 w^2 + \frac{1}{3!} m'''(x_{ij})b^3 w^3.
\] (5.19)

In a similar manner, the expansion of equation (5.17) up to order \(O(b^2)\) is given by:

\[
E \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} [m_1(x_{ij})] = \frac{1}{MN} \left\{ \begin{array}{c}
MN - mn \\
mnb
\end{array} \right\} \int k(w) (m'(x_{ij})bw + \frac{1}{2} m''(x_{ij}) b^2 w^2) \\
\times (g(x_{ij}) + g'(x_{ij})bw dw \}.
\] (5.20)

Simplifying equation (5.21) gives:

\[
E \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} [m_1(x_{ij})] = \frac{1}{MN} \left\{ \begin{array}{c}
(MN - mn) g(x_{ij}) m'(x_{ij}) b \int w k(w) dw \\
+ \left( \frac{MN - mn}{mn} \right) g'(x_{ij}) m'(x_{ij}) b^2 \int w^2 k(w) dw \\
+ \left( \frac{MN - mn}{mn} \right) \frac{1}{2} g(x_{ij}) m''(x_{ij}) b^2 \int w^2 k(w) dw + O(b^2) \right\}.
\] (5.21)

Using the conditions stated in equation (4.2), the derivation in equation (5.21) can further be simplified to obtain:

\[
E \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} [m_1(x_{ij})] = \frac{1}{MN} \left\{ \begin{array}{c}
(MN - mn) \left[ g'(x_{ij}) m'(x_{ij}) + \frac{1}{2} g(x_{ij}) m''(x_{ij}) \right] b^2 k \\
+ O(b^2) \right\}.
\] (5.22)
Hence the expected value of the second term in equation (5.12) then becomes:

\[
E \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} m_1(x_{ij}) = \frac{1}{MN} \left\{ \left( \frac{MN - mn}{mn} \right) \left[ \frac{1}{2} m'(x_{ij}) + \frac{g'(x_{ij}) m'(x_{ij})}{g(x_{ij})} \right] b^2 d_k + O(b^2) \right\}.
\]  

(5.23)

Simplifying equation (5.23) gives:

\[
E \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} m_1(x_{ij}) = \frac{1}{MN} \left\{ \left( \frac{MN - mn}{mn} \right) b^2 d_k C(x) + O(b^2) \right\}.
\]  

(5.24)

where \( C(x) = \frac{1}{2} m''(x_{ij}) + \left[ g(x_{ij}) \right]^{-1} g'(x_{ij}) m'(x_{ij}) \) and \( d_k \) is as stated in equation (4.2).

Using the bias given in (5.4) and the conditional expectation in equation (5.12), we obtain the following equation for the bias of the estimator:

\[
\text{Bias}(\hat{Y}) = \frac{1}{MN} \left\{ \left( \frac{MN - mn}{mn} \right) b^2 d_k C(x) + O(b^2) \right\}.
\]  

(5.25)

6. Asymptotic variance of the estimator of the population mean

From equation (5.9) and (5.12), we have

\[
m_2(x_{ij}) = \frac{1}{mnb} \sum_{i=1}^{n} \sum_{j=1}^{m} K(X_{ij} - x_{ij}) e_{ij}.
\]  

(6.1)

Hence

\[
\text{Var} \sum_{i=1}^{N} \sum_{j=1}^{M} m_2(x_{ij}) = \frac{1}{(MN)^2} \left( \frac{MN - mn}{mn} \right)^2 \sum_{i=1}^{n} \sum_{j=1}^{m} \text{Var}(D_x).
\]  

(6.2)

where \( D_x = K \left( \frac{X_{ij} - x_{ij}}{b} \right) e_{ij} \).

Expressing equation (6.2) in terms of expectation we obtain

\[
\text{Var} \sum_{i=1}^{N} \sum_{j=1}^{M} m_2(x_{ij}) = \frac{1}{(MN)^2} \left( \frac{(MN - mn)^2}{mn b^2} \right) \{ E[D_x]^2 - [E(D_x)]^2 \}.
\]  

(6.3)

Using the fact that the conditional expectation \( E(e_{ij}/X_{ij}) = 0 \), the second term in equation (6.3) can be reduced to zero. Therefore,

\[
\text{Var} \sum_{i=1}^{N} \sum_{j=1}^{M} m_2(x_{ij}) = \frac{1}{(MN)^2} \left( \frac{(MN - mn)^2}{mn b^2} \right) \sigma_{ij}^2.
\]  

(6.4)

where \( E(e_{ij}/X_{ij})^2 = \sigma_{ij}^2 \).

Let \( X = X_{ij} \) and \( x = x_{ij} \) and let \( w = \frac{X - x}{b} \), \( X - x = bw \) and differentiate \( X \) with respect to \( w \) as shown in equation (6.5)

\[
\begin{align*}
w &= \frac{X - x}{b} \\
X - x &= bw \\
dX &= bdw.
\end{align*}
\]  

(6.5)
so that
\[ Var \sum_{i=1}^{N} \sum_{j=1}^{M} [m_2(x_{ij})] = \frac{(MN - mn)^2}{mnb^2(MN)^2} \int K \left( \frac{X - x}{b} \right)^2 \sigma_x^2 g(X) dX. \] (6.6)

Equation (6.6) can also be written as
\[ Var \sum_{i=1}^{N} \sum_{j=1}^{M} [m_2(x_{ij})] = \frac{(MN - mn)^2}{mnb^2(MN)^2} \int K(w)^2 \sigma_x^2 g(x + bw) dw, \] (6.7)

which can be simplified to get
\[ Var \sum_{i=1}^{N} \sum_{j=1}^{M} [m_2(x_{ij})] = \frac{(MN - mn)^2}{mnb^2(MN)^2} \int K(w)^2 \sigma_x^2 g(x) dw + O \left( \frac{1}{mnb} \right). \] (6.8)

\[ Var \sum_{i=1}^{N} \sum_{j=1}^{M} [m_1(x_{ij})] \] can similarly be obtained as follows
\[ Var \sum_{i=1}^{N} \sum_{j=1}^{M} [m_1(x_{ij})] = \frac{1}{(MN)^2} Var \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} \left[ \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} K \left( \frac{X_{ij} - x_{ij}}{b} \right) \right] [M(X_{ij}) - m(x_{ij})]. \] (6.9)

Equation (6.9) can be rewritten as
\[ Var \sum_{i=1}^{N} \sum_{j=1}^{M} [m_1(x_{ij})] = \frac{(MN - mn)^2}{mnb^2(MN)^2} \int M(K)^2 [M(X_{ij}) - m(x_{ij})]^2 g(X) dX, \] (6.10)

where \( X = bw + x \) so that \( dX = bdw \). Changing variables and applying Taylor’s series expansion then
\[ Var \sum_{i=1}^{N} \sum_{j=1}^{M} [m_1(x_{ij})] = \frac{(MN - mn)^2}{mnb^2(MN)^2} \int K(w)^2 [m(x + bw) - m(x)]^2 g(x + bw) dw, \] (6.11)

which can be simplified to
\[ Var \sum_{i=1}^{N} \sum_{j=1}^{M} [m_1(x_{ij})] = \frac{(MN - mn)^2}{mnb^2(MN)^2} \int K(w)^2 [m(x) + m'(x)bw + \cdots - m(x)]^2 g(x) + g'(x)bw) dw. \] (6.12)

Simplifying equation (6.12) reduces to
\[ Var \sum_{i=1}^{N} \sum_{j=1}^{M} [m_1(x_{ij})] = O \left[ \frac{(MN - mn)^2b^2}{mnb} \right]. \] (6.13)

For large samples, as \( n \to N, m \to M \) and \( b \to 0 \), then \( mnb \to \infty \). Hence the variance in equation (6.12) asymptotically tends to zero, i.e, \( Var \sum_{i=1}^{N} \sum_{j=1}^{M} [m_1(x_{ij})] \to 0 \) so that the
variance of the estimator of the population mean reduces to

\[ \text{Var}(\hat{Y}) = \frac{(MN - mn)^2}{mnb(MN)^2} \sum_{i=n+1}^{M} \sum_{j=m+1}^{N} \text{Var} \left[ m(x_{ij}) + m'(x_{ij}) + m''(x_{ij}) \right]. \] (6.14)

Simplifying equation (6.14) we get

\[ \text{Var}(\hat{Y}) = \frac{(MN - mn)^2}{mnb(MN)^2} \sum_{i=n+1}^{N} \sum_{j=m+1}^{M} \left[ m(x_{ij}) \right]. \] (6.15)

Substituting equation (6.8) into (6.15) yields the following:

\[ \text{Var}(\hat{Y}) = \frac{1}{(MN)^2} \left\{ \frac{(MN - mn)^2}{mnb(g(x_{ij}))} \int K(w)^2 \sigma^2 x_{ij} dw + O \left[ \frac{(MN - mn)^2}{mnb} + \frac{1}{mn} \right] \right\}. \] (6.16)

This can further be simplified to get

\[ \text{Var}(\hat{Y}) = \frac{1}{(MN)^2} \left\{ \frac{(MN - mn)^2}{mnb(g(x_{ij}))} H(w)^2 \sigma^2 x_{ij} dw + O \left[ \frac{(MN - mn)^2}{mnb} + \frac{1}{mn} \right] \right\}. \] (6.17)

where \( H(w) = \int K(w)^2 dw \).

It is notable that the variance term still depends on the marginal density function, \( g(x_{ij}) \) of the auxiliary variables \( X_{ij} \). It can also be observed that the variance is inversely related to the smoothing parameter, \( b \). This implies that an increase in \( b \) results in a smaller variance. However, increasing the bandwidth would give a larger bias. Therefore there is a trade-off between the bias and the variance of the estimated population mean. A bandwidth that provides a compromise between the two measures would therefore be desirable.

7. Mean Squared Error (MSE) of the Estimator of Finite Population Mean

The MSE of the estimator combines the bias and the variance terms of estimator, \( \overline{\hat{Y}} \), that is,

\[ \text{MSE}(\overline{\hat{Y}}) = E(\overline{\hat{Y}} - \overline{Y})^2. \] (7.1)

Equation (7.1) can be re-written as:

\[ \text{MSE}(\overline{\hat{Y}}) = E(\overline{\hat{Y}} - E[\overline{\hat{Y}}] + E[\overline{\hat{Y}}] - \overline{Y})^2. \] (7.2)

Expanding equation (7.2) gives

\[ \text{MSE}(\overline{\hat{Y}}) = E(\overline{\hat{Y}} - E[\overline{\hat{Y}}])^2 + E(\text{Var}[\overline{\hat{Y}}]) + 2E(\text{Var}[\overline{\hat{Y}}])(\overline{\hat{Y}} - E[\overline{\hat{Y}}]). \] (7.3)

Equation (7.3) shows that

\[ \text{MSE}(\overline{\hat{Y}}) = \text{Var}(\overline{\hat{Y}}) + \text{Bias}^2 + 0. \] (7.4)
Combining the bias in equation (5.25) and the variance in equation (6.17) and conditioning on the auxiliary values $x_{ij}$ of the auxiliary variables $X_{ij}$, we get

$$MSE(\hat{Y}/X_{ij} = x_{ij}) = \frac{1}{(MN)^2} \left\{ \frac{(MN - mn)^2 H(w)\sigma^2_{x_{ij}}}{mnb(g(x_{ij}))} + O\left( \frac{1}{MN} \frac{(MN - mn)^2}{mnb} \right) \right\},$$

(7.5)

which can be simplified to

$$MSE(\hat{Y}/X_{ij} = x_{ij}) = \frac{1}{(MN)^2} \left\{ \frac{(MN - mn)^2 H(w)\sigma^2_{x_{ij}}}{mnb(g(x_{ij}))} + \frac{(MN - mn)^2}{4(\bar{m}n)^2(MN^2)} b^2 d_k^2 \left[ m''(x_{ij}) + \frac{2g'(x_{ij})m'(x_{ij})}{g(x_{ij})} \right]^2 \right\} + O\left( \frac{1}{MN} \frac{(MN - mn)^2}{mnb} + \frac{1}{mnb} \right).$$

(7.6)

Where $H(w) = \int K(w)^2 dw$, $d_k = \int w^2 K(w)dw$, $C(x) = \frac{1}{2}m''(x_{ij}) + [g(x_{ij})]^{-1}g'(x_{ij})m'(x_{ij})$ as used earlier in the rest of the derivations.

8. Conclusion

If the sample size is large, that is as $n \to N$ and $m \to M$, the $MSE$ of $\hat{Y}$ in equation (7.6) tends to zero for a sufficiently small bandwidth, $b$. The estimator $\hat{Y}$ is therefore asymptotically consistent since its $MSE$ converges to zero.

References

[1] Bethlehem J 2012 Using response probabilities for assessing representativity Statistics Netherlands, Discussion Paper.
[2] Dorfman A H 1992 Nonparametric regression for estimating totals in finite populations Proceedings of the Section on Survey Research Methods 622-625.
[3] Nadaraya E A 1964 On estimating regression Theory of Probability and Its Applications 9(1)141–142.
[4] Onyango C O, Otieno R O and Orwa G O 2010 Generalised Model Based Confidence Intervals in Two Stage Cluster Sampling Pakistan Journal of Statistics and Operation Research 6(2).
[5] Ouma C and Wafula C 2005 Bootstrap confidence intervals for model-based surveys East African Journal of Statistics, JKUAT Press Ltd 1(1)84–90.
[6] Lee H, Rancourt E and Särndal C E 2002 Variance estimation from survey data under single imputation New York, Survey nonresponse 315–328.
[7] Singh S and Horn S 2000 Compromised imputation in survey sampling Springer, Metrika 51(3)267–276.
[8] Watson G S 1964 Smooth regression analysis JSTOR, Sankhýā: The Indian Journal of Statistics, Series A359–372.