

## Estimation of Response Bias from Both Samples in Double Sampling with Unequal Probabilities

Godwin A. Udofia  
University of Calabar, Cross River State, Southeaster, Nigeria

**Abstract:** In this study, we consider estimation of response bias based on information from both the initial and the subsample in double sampling with inclusion probabilities in contrast to estimation of the response bias based on information from the subsample only. The condition, under which either of the two estimators can be preferred to the other is determined. The relative significance of the contribution of the correlated factors to the total response variance compared to the contribution by the other factors of response variation noted earlier by Hansen under a different sample design is confirmed. The effect of the replication of the survey for  $K > 1$  times.

**Key words:** Response bias, both samples, double sampling, unequal probabilities, comparison of estimators

---

### INTRODUCTION

The term survey is used in this study to refer to an investigation, whether complete or incomplete, which is concerned mainly with the counting of units (of analysis) under study, technically known as a census (Warwick and Lininger, 1975; Raj, 1968) and to a complete or incomplete investigation, which is concerned with the collection of detailed information, technically known as a survey (Warwick and Lininger, 1975), from each unit under study. We postulate that the objective of a statistician in designing a survey is to minimize the total survey error the components of which can be classified broadly, by error type, as sampling errors and nonsampling errors.

The sources of survey errors are all the activities that constitute survey planning or determination of the fixed conditions (though often ignored as a source of error) under which, the survey can be conducted. These conditions include population coverage and the survey protocol (Udofia, 2006). The planning operations take place in the office. The other source of survey errors is the process of measurement of the study variable on the units of analysis. This process consists of Hansen *et al.* (1961) the data collection operations (or field work) and the data processing operations (which is another phase of office work).

Sampling error arises from failure of the selected (random) sample to cover the entire population. It is therefore, associated with certain aspects of the survey planning, which are choice of sample design and determination of sample size. This error is a variable the magnitude of which is inversely related to the sample size.

This relationship enables the investigator to reduce the sampling error by increasing the sample size. The close relationship between this class of survey errors and the sample size gives rise to the term sampling error. Nonsampling errors are associated with the other aspects of planning (such as the selection and training of the survey staff and the survey time schedule), with the measurement and with the processing operations.

Non-sampling errors, the reduction of which is receiving increasing attention in recent times, are different distortions of the true value of the study variable. They consist of biases or systematic distortions and random errors both of which are referred to commonly as response errors.

Random errors appear in equal magnitudes in both positive and negative directions and hence cancel out in summation. Biases on the other hand, are systematic distortions which pile up in the same direction and therefore do not cancel out in the summation. As indicated above, these errors are not functions of the sample size. They remain the same in both complete and in incomplete surveys. As noted by Kish (1965) and by Udofia (2008), they can be reduced only by improving the related survey operations in the planning, measurement and in the data processing stages.

It is known (Kish, 1965 and Talukder, 1975) that in many survey situations, response bias is the major component of response error. Raj (1965) notes that sometimes it is so large that its elimination is more important than increasing the sample size. For example, Hansen *et al.* (1961) concludes that response bias was the important contribution to the errors of the 1950 U.S. Census statistics.

Measurement of response bias is a difficult problem, the solution of which demands more than statistical theory (Udofia, 2008). Hansen *et al.* (1961) concludes that there is no reasonably satisfactory approach for measurement of response bias. After the pioneer research by Mahalanobis (1946) on measurement of response variation some useful methods of measurement of response bias have been developed and discussed by notably Hansen *et al.* (1951, 1953, 1961), Fellegi (1964), Raj (1968), Koch (1973) and Talukder (1975). Hansen *et al.* (1951) gives an illustration of the use of double sampling with equal probabilities of selection in each phase to measure response bias. It is of interest to see how information from both the initial sample and the 2nd sample can be used together for the same purpose and to determine the condition under which either estimator can be preferred. This study is an attempt in this direction.

Suppose that the population,  $\pi_1$ , under study consists of  $N$  elements which can be divided into  $H$  identifiable and independent groups or subpopulations, which may be geographic, demographic or professional groups. Let,

$$N_h, \sum_{h=1}^H N_h = N$$

denote the number of potential respondents or units of analysis in the  $h$ th group ( $G_h$ ). There exists also a population,  $\pi_2$  of  $M$  potential interviewers which can be divided also into  $H$  independent groups with  $M_h$ ,

$$\sum_{h=1}^H M_h = M$$

potential interviewers in the  $h$ th group ( $Q_h$ ) such that units in  $G_h$  can be interviewed only by units in  $Q_h$ . As noted by Hansen *et al.* (1951), the discussion in this study applies also to a situation where the elements of  $\pi_2$  can be divided into  $L$ ,  $L \neq H$ , subpopulations. An example of the arrangement described above is given in Udofia (2008). The number of potential respondents in  $G_h$  that are available to be interviewed by each of the  $M_h$  units of  $Q_h$  is

$$\bar{N}_h = \left[ \frac{N_h}{M_h} \right]$$

It is assumed that information on the measure of size,  $Z$ , is available for each of the population elements,  $U_j$   $j = 1, 2, \dots, N$ , before the start of the survey. An initial sample,  $S(n_{1h})$  of size  $n_{1h}$ ,

$$\sum_{h=1}^H n_{1h} = n_1$$

is drawn from  $G_h$ ,  $h = 1, 2, \dots, H$ , with probability

$$P_{hj} = Z_{hj} / Z_h, \quad Z_h = \sum_{j=1}^{N_h} Z_{hj}$$

where,  $Z_{hj}$  is the value of  $Z$  for unit  $j$  and with replacement. Also, a sample  $S(m_{1h})$  of  $m_{1h}$

$$\sum_{j=1}^{m_{1h}} m_{1h} = m_1$$

interviewers is drawn by Simple Random Sampling Without Replacement (SRSWOR) from  $Q_h$ ,  $h = 1, 2, \dots, H$  and

$$\bar{n}_{1h} = \left[ \frac{n_{1h}}{m_{1h}} \right]$$

units of  $S(n_{1h})$  are assigned at random to each of the  $m_{1h}$  interviewers so that  $n_{1h} = \bar{n}_{1h} m_{1h}$ .

The study variable is measured on  $S(n_{1h})$  under a set,  $A$ , of essential survey conditions (Udofia, 2006) typical of a large survey. Let,  $X_{hijk}$  denote the value of the study variable for unit  $j$  in  $G_h$  obtained by interviewer  $i$  in  $Q_h$  on the  $k$ th,  $k = 1, 2, \dots, K$  visit under this set of essential survey conditions. This means that a survey can be repeated  $K$ ,  $K > 1$ , times (Hansen *et al.*, 1951; Raj, 1968). The repetition can be done in such a way that information received from an individual on one visit does not influence information received from that individual on any subsequent visit.

During the second phase of the survey, a second sample,  $S(n_{2h})$  of size  $n_{2h}$ ,

$$\sum_{h=1}^H n_{2h} = n_2$$

$n_{2h} < n_{1h}$ , is drawn by SRSWOR from  $S(n_{1h})$ . Also, a sample  $S(m_{2h})$  of  $m_{2h}$

$$m_{2h} < m_{1h}, \quad \sum_{h=1}^H m_{2h} = m_2$$

interviewers is drawn by SRSWOR from the initial  $m_{1h}$  interviewers (or from another population of interviewers that is different from  $\pi_2$ ) and

$$\bar{n}_{2h} = \left[ \frac{n_{2h}}{m_{2h}} \right]$$

units of  $S(n_{2h})$  are assigned at random to each of the  $m_{2h}$  interviewers. The study variable is measured once on  $S(n_{2h})$  under a set,  $B$ , of highly effective essential survey conditions that can produce only the desired true or precise responses (Udofia, 2006). Let,  $Y_{hij}$  denote the value of the study variable obtained by interviewer  $i$ ,  $i \in S(m_{2h})$  from unit  $j$ ,  $j \in S(n_{2h})$  under set  $B$  of essential survey conditions. These are the desired true values.

Consider the following linear response model of the type discussed earlier by Hansen *et al.* (1953), Raj (1968) and by Talukder (1975).

$$X_{hijk} = Y_{hij} + \alpha_{hi} + \beta_{hj} + d_{hijk} \quad (1)$$

where:

$X_{hijk}$  and  $Y_{hij}$  = Retain their earlier definitions  
 $\alpha_{hi}$  = Denotes bias due to interviewer  $i$   
 $i \in Q_h, \beta_{hj}$  = Denotes bias due to respondent  $j, j \in G_h$   
 $d_{hijk}$  = Denotes random error in the response by unit  $j$  received by interviewer  $i$  on occasion  $k$

We assume that

$$d_{hijk} = 0, \text{ var}(d_{hijk}) = \sigma^2(d) \quad (2)$$

$$\text{Cov}(d_{hijk}, d_{hiwt}) = \begin{cases} 0; i = \ell, j = w, t \neq k \\ 0; i \neq \ell, j \neq w, t = k \\ 0; i \neq \ell, j \neq w, t \neq k \\ \rho(d_{hijk}, d_{hiwt})\sigma^2(d); i = \ell, j \neq w, t = k \\ \sigma^2(d); i = \ell, j = w, t = k \\ 0; i = \ell, j \neq w, t \neq k \end{cases}$$

We assume also that there is no correlation between elements of the set  $\{Y_{hij}, \alpha_{hi}, \beta_{hj}, d_{hijk}\}$ . Under these assumptions, response bias is defined as:

$$B = \sum_{h=1}^H \sum_{i=1}^{M_h} \sum_{j=1}^{N_h} \sum_{k=1}^K X_{hijk} - \sum_{h=1}^H \sum_{i=1}^{M_h} \sum_{j=1}^{N_h} Y_{hij} \quad (3)$$

$$= \sum_{h=1}^H \bar{N}_h \sum_{j=1}^{M_h} \alpha_{hi} + \sum_{h=1}^H M_h \sum_{j=1}^{N_h} \beta_{hj}$$

or

$$B = \sum_{h=1}^H N_h (\bar{X}_h - \bar{Y}_h) = X - Y = \sum_{h=1}^H N_h (\bar{\alpha}_h + \bar{\beta}_h) \quad (4)$$

since,

$$\sum_{h=1}^H d_{hijk} = 0$$

by (2). Equation (3) clearly presents total bias,  $B$ , as the sum of all biases (Kish, 1965).

Under the above assumptions, we consider the following unbiased estimator of  $B$  based on information on the subsample,  $S(n_{2h})$  only:

$$\hat{B}^* = \sum_{h=1}^H \frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} - \sum_{h=1}^H \frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hij}}{P_{hj}}; K > 1 \quad (5)$$

with variance

$$V(\hat{B}^*) = \sum_{h=1}^H \frac{K}{n_{1h}} \sum_{j=1}^{M_h} \sigma_p^2(X)_{hi} \left[ 1 + (\bar{n}_{1h} - 1) \rho_p(X_{hijk}, X_{hi\ell k})_{hi} \right. \\ \left. + (K-1) \rho_p(X_{hijk}, X_{hiw})_{hi} \right] + \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{j=1}^{M_h} \sigma_p^2(Y)_{hi} \left[ 1 + (\bar{n}_{1h} - 1) \rho_p \right. \\ \left. (Y_{hij}, Y_{hi\ell})_{hi} \frac{\sigma_p(X)_{hi}}{\sigma_p(Y)_{hi}} \right] + \sum_{h=1}^H \left( \frac{1}{n_{2h}} - \frac{1}{n_{1h}} \right) \\ \sum_{j=1}^{M_h} \left[ \sigma_p^2(X)_{hi} + \sigma_p^2(Y)_{hi} - 2 \rho_p(X_{hijk}, Y_{hij})_{hi} \sigma_p(X)_{hi} \sigma_p(Y)_{hi} \right] \quad (6)$$

This variance can also be expressed in terms of its simple components that reflect the different sources of response variation as:

$$V(\hat{B}^*) = \sum_{h=1}^H \frac{K}{n_{1h}} \sum_{j=1}^{M_h} \left\{ \sigma_p^2(d)_{hi} \left[ 1 + (\bar{n}_{1h} - 1) \rho_p(d_{hijk}, d_{hi\ell k})_{hi} \right] \right. \\ \left. + (\bar{n}_{1h} - 1) \rho_p(Y_{hij}, Y_{hi\ell})_{hi} \sigma_p^2(Y)_{hi} + \left[ K + (\bar{n}_{1h} - 1) \rho_p(\beta_{hj}, \beta_{hi\ell})_{hi} \right] \right. \\ \left. \frac{\sigma_p^2(\beta)_{hi}}{M_h} + \frac{\sigma_p^2(\alpha)_{hi}}{M_h} \right\} + \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{j=1}^{M_h} \left[ (K-1) \sigma_p^2(Y)_{hi} + (\bar{n}_{1h} - 1) \rho_p \right. \\ \left. (Y_{hij}, Y_{hi\ell})_{hi} \sigma_p^2(Y)_{hi} \right] + \sum_{h=1}^H \left( \frac{1}{n_{2h}} - \frac{1}{n_{1h}} \right) \sum_{j=1}^{M_h} \left[ (K-1) \sigma_p^2(Y)_{hi} \right. \\ \left. + K \bar{N}_h \sigma^2(\alpha)_{hi} + \sigma^2(d)_{hi} + K \sigma_p^2(\beta)_{hi} \right] \quad (7)$$

Contributions by the correlated factors are represented in Eq. 7 by the terms with  $(\bar{n}_{1h} - 1)$  as a factor and which, as noted by Hansen *et al.* (1961), can be large.

## METHOD OF ESTIMATION OF RESPONSE BIAS FROM BOTH SAMPLES IN DOUBLE SAMPLING WITH UNEQUAL PROBABILITIES

Under the sample design and with the data in previous study, we propose the following alternative estimator of response bias to include information from both phases of the double sampling design:

$$\hat{B} = \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} - \sum_{h=1}^H \frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hij}}{P_{hj}} \quad (8)$$

By using theorems on conditional expectations and conditional variances (Raj, 1956, 1968), we obtain the expected value and the variance of the above estimator as follows:

By the conditional expectation equation:

$$E(\hat{B}) = E_1[E_2(\hat{B})]; E_\lambda(\cdot) = E(\cdot | n_\lambda) \quad (9)$$

Where:

$$E_2(\hat{B}) = E_2 \left[ \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} - \sum_{h=1}^H \frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hij}}{P_{hj}} \right] \quad (10)$$

$$= \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} - \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \frac{Y_{hij}}{P_{hj}}$$

and hence,

$$E(\hat{B}) = E_1 \{ E_2(\hat{B}) \}$$

$$= \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \left( \sum_{k=1}^K \sum_{l=1}^{\bar{N}_h} P_{hj} \frac{X_{hijk}}{P_{hj}} \right) - \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \left( \sum_{k=1}^K \sum_{l=1}^{\bar{N}_h} P_{hj} \frac{Y_{hij}}{P_{hj}} \right)$$

$$= \sum_{h=1}^H \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \sum_{k=1}^K X_{hijk} - \sum_{h=1}^H \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} Y_{hij} = \sum_{h=1}^H \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} (X_{hij} - Y_{hij}) = X - Y$$

which is the same as Eq. (4) where,

$$X_{hij} = \sum_{k=1}^K X_{hijk}$$

The variance of  $\hat{B}$  is given by the conditional variance equation:

$$V(\hat{B}) = V_1 E_2(\hat{B}) + E_1 V_2(\hat{B}); \quad V_\lambda(\cdot) = V(\cdot | n_\lambda) \quad (11)$$

From Eq. (10), we obtain

$$V_1 E_2(\hat{B}) = V_1 \left( \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} \right) + V_1 \left( \sum_{h=1}^H \frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hij}}{P_{hj}} \right)$$

$$- 2C_1 \left( \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}}, \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \frac{Y_{hij}}{P_{hj}} \right) \quad (12)$$

Now,

$$V_1 \left( \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} \right) = \sum_{h=1}^H \frac{K}{n_{1h}} V_1 \left( \frac{X_{hijk}}{P_{hj}} \right)_h$$

$$+ \sum_{h=1}^H \frac{K(\bar{n}_{1h} - 1)}{n_{1h}} C_1 \left( \frac{X_{hijk}}{P_{hj}}, \frac{X_{hi\ell k}}{P_{hi}} \right)_h$$

$$+ \sum_{h=1}^H \frac{K(K-1)}{n_{1h}} C_1 \left( \frac{X_{hijk}}{P_{hj}}, \frac{X_{hijw}}{P_{hi}} \right)_h$$

$$+ \frac{1}{2} \sum_{h=1}^H \frac{\bar{n}_{1h} - 1}{n_{1h}} K(K-1) C_1 \left( \frac{X_{hijk}}{P_{hj}}, \frac{X_{hijw}}{P_{hi}} \right)_h \quad (13)$$

We rewrite Eq. (13) as:

$$\sum_{h=1}^H \frac{K}{n_{1h}} \sum_{i=1}^{M_h} \sigma_p^2(X)_{hi} + \sum_{h=1}^H \frac{K(\bar{n}_{1h} - 1)}{n_{1h}} \sum_{i=1}^{M_h} \rho_p(X_{hijk}, X_{hi\ell k})_{hi} \sigma_p^2(X)_{hi}$$

$$+ \sum_{h=1}^H \frac{K(K-1)}{n_{1h}} \sum_{i=1}^{M_h} \rho_p(X_{hijk}, X_{hijw})_{hi} \sigma_p^2(X)_{hi} \quad (14)$$

$$= \sum_{h=1}^H \frac{K}{n_{1h}} \sum_{i=1}^{M_h} \sigma_p^2(X)_{hi} \left[ 1 + (\bar{n}_{1h} - 1) \rho_p(X_{hijk}, X_{hi\ell k})_{hi} \right]$$

$$+ (K-1) \rho_p(X_{hijk}, X_{hijw})_{hi}$$

Where:

$$\sigma_p^2(X)_{hi} = \sum_{j=1}^{\bar{N}_h} \sum_{k=1}^K P_{hj} \left( \frac{X_{hijk}}{P_{hj}} - X_{hi} \right)^2 = V \left( \frac{X_{hijk}}{P_{hj}} \right)_{hi}$$

$$C_1 \left( \frac{X_{hijk}}{P_{hj}}, \frac{X_{hi\ell k}}{P_{hi}} \right) = C_p(X_{hijk}, X_{hi\ell k})_{hi}$$

$$= \sum_{j \neq \ell} \sum_{k=1}^K P_{hj} \cdot P_{hi} \left( \frac{X_{hijk}}{P_{hj}} - X_{hi} \right) \left( \frac{X_{hi\ell k}}{P_{hi}} - X_{hi} \right)$$

$$C_1 \left( \frac{X_{hijk}}{P_{hj}}, \frac{X_{hijw}}{P_{hi}} \right) = 0; \quad j \neq \ell, \quad k \neq w \quad (15)$$

$$V_1 \left( \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \frac{Y_{hij}}{P_{hj}} \right) = \sum_{h=1}^H \frac{1}{n_{1h}} V_1 \left( \frac{Y_{hij}}{P_{hj}} \right)$$

$$+ \sum_{h=1}^H \frac{\bar{n}_{1h} - 1}{n_{1h}} C_1 \left( \frac{Y_{hij}}{P_{hj}}, \frac{Y_{hi\ell}}{P_{hi}} \right), \quad j \neq \ell = \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{M_h} \sigma_p^2(Y)_{hi}$$

$$\left[ 1 + (\bar{n}_{1h} - 1) \rho_p(Y_{hij}, Y_{hi\ell})_{hi} \right]$$

Where:

$$\sigma_p^2(Y)_{hi} = \sum_{j=1}^{\bar{N}_h} P_{hj} \left( \frac{Y_{hij}}{P_{hj}} - Y_{hi} \right)^2$$

Also,

$$C_1 \left( \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}}, \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \frac{Y_{hij}}{P_{hj}} \right)$$

$$= \sum_{h=1}^H \frac{1}{n_{1h}} C_1 \left( \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}}, \frac{Y_{hij}}{P_{hi}} \right)_h$$

Where:

$$C_1 \left( \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}}, \frac{Y_{hij}}{P_{hi}} \right)_h$$

$$= \sum_{j=1}^{M_h} \left[ \sum_{k=1}^{\bar{N}_h} \sum_{l=1}^K P_{hj} \left( \frac{X_{hijk}}{P_{hj}} - X_{hi} \right) \left( \frac{Y_{hij}}{P_{hi}} - Y_{hi} \right) \right]_{hi}$$

$$= \sum_{j=1}^{M_h} C_p(X_{hijk}, Y_{hij})_{hi}$$

so that

$$C_1 \left( \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}}, \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \frac{Y_{hij}}{P_{hj}} \right) \quad (16)$$

$$= \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{j=1}^{M_h} \rho_p(X_{hijk}, Y_{hij})_{hi} \sigma_p^2(X)_{hi} \sigma_p(Y)_{hi}$$

Substitution of Eq. (14-16) in Eq. (12) gives the result

$$V_1 E_2(\hat{B}) = \sum_{h=1}^H \frac{K}{n_{1h}} \sum_{i=1}^{M_h} \sigma_p^2(X)_{hi} \left[ 1 + (\bar{n}_{1h} - 1) \rho_p(X_{hijk}, X_{hi\ell k})_{hi} \right. \\ \left. + (K-1) \rho_p(X_{hijk}, X_{hijw})_{hi} \right] + \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{M_h} \sigma_p^2(Y)_{hi} \times \\ \times \left[ 1 + (\bar{n}_{1h} - 1) \rho_p(Y_{hij}, Y_{hi\ell})_{hi} - 2 \rho_p(Y_{hijk}, Y_{hi\ell})_{hi} \frac{\sigma_p(X)_{hi}}{\sigma_p(Y)_{hi}} \right] \quad (17)$$

From Eq. (8):

$$E_1 V_2(\hat{B}) = E_1 V_2 \left( \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} \right) \\ + E_1 V_2 \left( \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hij}}{P_{hj}} \right) \\ - 2 E_1 C_2 \left( \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}}, \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hij}}{P_{hj}} \right) \\ = E_1 V_2 \left( \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hij}}{P_{hj}} \right) = \sum_{h=1}^H E_1 V_2 \left( \frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hij}}{P_{hj}} \right)$$

Let,

$$\frac{Y_{hij}}{P_{hj}} = t_y$$

so that

$$\frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hij}}{P_{hj}} = \bar{t}_{2y}$$

Then

$$V_2(\hat{B}) = \sum_{h=1}^H V_2(\bar{t}_{2y}) \\ = \sum_{h=1}^H \left( \frac{1}{n_{2h}} - \frac{1}{n_{1h}} \right) \frac{1}{n_{1h} - 1} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \left( t_y - \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} t_y \right)^2 \\ = \sum_{h=1}^H \left( \frac{1}{n_{2h}} - \frac{1}{n_{1h}} \right) \frac{1}{n_{1h} - 1} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \left( \frac{Y_{hij}}{P_{hj}} - \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} t_y \right)^2$$

and

$$E_1 V_2(\hat{B}) = \sum_{h=1}^H \left( \frac{1}{n_{2h}} - \frac{1}{n_{1h}} \right) \sum_{i=1}^{M_h} \sum_{j=1}^{\bar{N}_h} P_{hj} \left( \frac{Y_{hij}}{P_{hj}} - Y_{hi} \right)_{hi}^2 \quad (18) \\ = \sum_{h=1}^H \left( \frac{1}{n_{2h}} - \frac{1}{n_{1h}} \right) \sum_{i=1}^{M_h} \sigma_p^2(Y)_{hi}$$

Substitution of Eq. (17) and (18) in Eq. (11) gives the result

$$V(\hat{B}) = \sum_{h=1}^H \frac{K}{n_{1h}} \sum_{i=1}^{M_h} \sigma_p^2(X)_{hi} \left[ 1 + (\bar{n}_{1h} - 1) \rho_p(X_{hijk}, X_{hi\ell k})_{hi} \right. \\ \left. + (K-1) \rho_p(X_{hijk}, X_{hijw})_{hi} \right] + \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{j=1}^{M_h} \sigma_p^2(Y)_{hi} \\ \left[ 1 + (\bar{n}_{1h} - 1) \rho_p(Y_{hij}, Y_{hi\ell})_{hi} - 2 \rho_p(X_{hijk}, Y_{hij})_{hi} \frac{\sigma_p(X)_{hi}}{\sigma_p(Y)_{hi}} \right] \\ + \sum_{h=1}^H \left( \frac{1}{n_{2h}} - \frac{1}{n_{1h}} \right) \sum_{j=1}^{M_h} \sigma_p^2(Y)_{hi} \quad (19)$$

The significant role of the correlated factors which are terms with factors  $\bar{n}_{1h} - 1$  and  $K-1$ , respectively is also obvious in this result.

For the purpose of decomposition of the terms in Eq. (13), we have under the model assumptions,

$$\sum_{h=1}^H \frac{K}{n_{1h}} V_1 \left( \frac{X_{hijk}}{P_{hj}} \right) = \sum_{h=1}^H \frac{K}{n_{1h}} V_1 \left( \frac{d_{hijk}}{P_{hj}} + \frac{Y_{hij}}{P_{hj}} + \frac{\beta_{hj}}{P_{hj}} + a_{hi} \right) \\ = \sum_{h=1}^H \frac{K}{n_{1h}} \left[ V_1 \left( \frac{d_{hijk}}{P_{hj}} \right) + V_1 \left( \frac{Y_{hij}}{P_{hj}} \right) + V_1 \left( \frac{\beta_{hj}}{P_{hj}} \right) + V_1(a_{hi}) \right] \\ = \sum_{h=1}^H \frac{K}{n_{1h}} \sum_{i=1}^{M_h} \left[ \sigma_p^2(d)_{hi} + \sigma_p^2(Y)_{hi} + \frac{\sigma_p^2(\beta)_h + \sigma^2(\alpha)_h}{M_h} \right] \quad (20)$$

$$\sum_{h=1}^H \frac{K(\bar{n}_{1h} - 1)}{n_{1h}} C_1 \left( \frac{X_{hijk}}{P_{hj}}, \frac{X_{hi\ell k}}{P_{hj}} \right)_h \\ = \sum_{h=1}^H \frac{K(\bar{n}_{1h} - 1)}{n_{1h}} \sum_{i=1}^{M_h} \left[ \rho_p(Y_{hij}, Y_{hi\ell})_{hi} \sigma_p^2(Y)_{hi} \right. \\ \left. + \rho_p(d_{hijk}, d_{hi\ell k})_{hi} \sigma_p^2(d)_{hi} \right. \\ \left. + \rho_p(\beta_{hj}, \beta_{hi})_{hi} \sigma_p^2(Y)_{hi} \frac{\sigma_p^2(\beta)_h}{M_h} \right] \quad (21)$$

$$\begin{aligned} & \sum_{h=1}^H \frac{K(K-1)}{n_{1h}} C_1 \left( \frac{X_{hijk}}{P_{hj}}, \frac{X_{hijw}}{P_{hj}} \right) \\ &= \sum_{h=1}^H \frac{K(K-1)}{n_{1h}} \sum_{i=1}^{M_h} \left[ \sigma_p^2(Y)_{hi} + \frac{\sigma_p^2(\beta)_h}{M_h} \right. \\ & \quad \left. + \rho_p(d_{hijk}, d_{hijw})_{hi} \sigma_p^2(d)_{hi} \right] \quad (22) \\ &= \sum_{h=1}^H \frac{K(K-1)}{n_{1h}} \sum_{i=1}^{M_h} \left[ \sigma_p^2(Y)_{hi} + \frac{1}{M_h} \sigma_p^2(\beta)_h \right] \end{aligned}$$

Under the model assumptions (2),  $\rho_p(d_{hijk}, d_{hijw})_{hi} = 0$ , for all  $k \neq w$ . Also, for any given  $J$ ,  $\sigma_p^2(Y)_{hi} = 0$  for all  $k \neq w$ , since,  $Y_{hij}$  is the same for all  $k \neq w$  under fixed  $i$  and  $j$ .

$$C_1 \left( \frac{X_{hijk}}{P_{hj}}, \frac{X_{hijw}}{P_{hj}} \right) = 0; j \neq \ell, k \neq w \quad (23)$$

Substitution of Eq. 20-23 in Eq. 13 gives, under the above assumptions, the following result.

$$\begin{aligned} & V_1 \left( \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} \right) \\ &= \sum_{h=1}^H \frac{K}{n_{1h}} \sum_{i=1}^{M_h} \left\{ \sigma_p^2(d)_{hi} \left[ 1 + (\bar{n}_{1h} - 1) \rho_p(d_{hijk}, d_{hi\ell k})_{hi} \right] \right. \\ & \quad + \sigma_p^2(Y)_{hi} \left[ 1 + (\bar{n}_{1h} - 1) \rho_p(Y_{hij}, Y_{hi\ell})_{hi} \right] \\ & \quad + \frac{\sigma_p^2(\beta)_h}{M_h} \left[ K + (\bar{n}_{1h} - 1) \rho_p(Y_{hij}, Y_{hi\ell})_{hi} \rho_p(\beta_{hij}, \beta_{hi\ell})_h \right] \\ & \quad \left. + \frac{\sigma_p^2(\alpha)_h}{M_h} \right\} \quad (24) \end{aligned}$$

The 2nd term in Eq. 12 retains its definition in Eq. 15. The last term in Eq. 12, when  $X_{hijk}$  is expressed in terms of its components as in Eq. (1), becomes:

$$\begin{aligned} & C_1 \left( \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}}, \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{m_{1h}} \sum_{j=1}^{\bar{n}_{1h}} \sum_{k=1}^K \frac{Y_{hij}}{P_{hj}} \right) \\ &= \sum_{h=1}^H \frac{1}{n_{1h}} C_1 \left( \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}}, \frac{Y_{hij}}{P_{hj}} \right) = \\ & \sum_{h=1}^H \frac{1}{n_{1h}} C_1 \left( \frac{Y_{hij}}{P_{hj}} + \frac{\beta_{hij}}{P_{hj}} + \alpha_{hi}, \frac{Y_{hij}}{P_{hj}} \right) \quad (25) \\ &= \sum_{h=1}^H \frac{1}{n_{1h}} C_1 \left( \frac{Y_{hij}}{P_{hj}} + \frac{Y_{hij}}{P_{hj}} \right) = \sum_{h=1}^H \frac{1}{n_{1h}} V_p(Y)_h \\ &= \sum_{h=1}^H \frac{1}{n_{1h}} \sum_{i=1}^{M_h} \sigma_p^2(Y)_{hi} \end{aligned}$$

Substitution of Eq. 15, 24 and 25 in Eq. 12 gives the result

$$\begin{aligned} & V_1 E_2(\hat{B}) = \sum_{h=1}^H \frac{K}{n_{1h}} \sum_{i=1}^{M_h} \left\{ \sigma_p^2(d)_{hi} \left[ 1 + (\bar{n}_{1h} - 1) \rho_p(d_{hijk}, d_{hi\ell k})_{hi} \right] \right. \\ & \quad + (\bar{n}_{1h} - 1) \rho_p(Y_{hij}, Y_{hi\ell})_{hi} \sigma_p^2(Y)_{hi} \\ & \quad + \frac{\sigma_p^2(\beta)_h}{M_h} \left[ K + (\bar{n}_{1h} - 1) \rho_p(\beta_{hij}, \beta_{hi\ell})_h \right] \\ & \quad + \frac{1}{M_h} \sigma_p^2(\alpha)_h \left. \right\} + \sum_{h=1}^H \frac{K}{n_{1h}} \sum_{i=1}^{M_h} \left[ (K-1) \sigma_p^2(Y)_{hi} \right. \\ & \quad \left. + (\bar{n}_{1h} - 1) \rho_p(Y_{hij}, Y_{hi\ell})_{hi} \sigma_p^2(Y)_{hi} \right] \quad (26) \end{aligned}$$

Now  $V_1 E_2(\hat{B})$  remains the same as in Eq. 18. The substitution of Eq. 18 and 26 in 11 gives the following expression for  $V(\hat{B})$  terms of simple response variances.

$$\begin{aligned} & V(\hat{B}) = \sum_{h=1}^H \frac{K}{n_{1h}} \sum_{i=1}^{M_h} \left\{ \sigma_p^2(d)_{hi} \left[ 1 + (\bar{n}_{1h} - 1) \rho_p(d_{hijk}, d_{hi\ell k})_{hi} \right] \right. \\ & \quad + (\bar{n}_{1h} - 1) \rho_p(Y_{hij}, Y_{hi\ell})_{hi} \sigma_p^2(Y)_{hi} \\ & \quad + \frac{\sigma_p^2(\beta)_h}{M_h} \left[ K + (\bar{n}_{1h} - 1) \rho_p(\beta_{hij}, \beta_{hi\ell})_h \right] + \frac{1}{M_h} \sigma_p^2(\alpha)_h \left. \right\} \\ & \quad + \sum_{h=1}^H \frac{K}{n_{1h}} \sum_{i=1}^{M_h} \left[ (K-1) \sigma_p^2(Y)_{hi} + (\bar{n}_{1h} - 1) \rho_p(Y_{hij}, Y_{hi\ell})_{hi} \sigma_p^2(Y)_{hi} \right] \\ & \quad + \sum_{h=1}^H \left( \frac{1}{n_{2h}} - \frac{1}{n_{1h}} \right) \sum_{i=1}^{M_h} \sigma_p^2(Y)_{hi} \quad (27) \end{aligned}$$

The terms on the right-hand-side of the above equation, which we refer to as the simple response variances, reflect contributions from the main sources of response variation which are the biases, variation of the true values and the correlated factors.

With regards to unbiased estimation of  $V(\hat{B})$  in Eq. 19, we give the following similarly to Raj (1968):

$$\begin{aligned} & E \left[ \frac{1}{Kn_{2h} - 1} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \left( \frac{X_{hijk}}{P_{hj}} - \frac{1}{Kn_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} \right)^2 \right] = \sigma_p^2(X)_{hi} \\ & \text{or} \\ & E \left[ \sum_{i=1}^{m_{2h}} \frac{1}{Kn_{2h} - 1} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \left( \frac{X_{hijk}}{P_{hj}} - \frac{1}{Kn_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} \right)^2 \right] = \sum_{i=1}^{M_h} \sigma_p^2(X)_{hi} \\ & \text{or} \\ & E \left[ \frac{1}{Kn_{2h} - 1} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \left( \frac{X_{hijk}}{P_{hj}} - \frac{1}{Kn_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} \right) \right. \\ & \quad \times \left( \frac{X_{hi\ell k}}{P_{h\ell}} - \frac{1}{Kn_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \frac{X_{hi\ell k}}{P_{h\ell}} \right) \left. \right] \\ &= \rho_p(X_{hijk}, X_{hi\ell k})_{hi} \sigma_p^2(X)_h \end{aligned}$$

or

$$\begin{aligned} & E \left[ \sum_{i=1}^{m_{2h}} \frac{1}{K\bar{n}_{2h}-1} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \left( \frac{X_{hijk}}{P_{hj}} - \frac{1}{K\bar{n}_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} \right) \right]_{hi} \\ & \times \left( \frac{X_{hi\ell k}}{P_{h\ell}} - \frac{1}{K\bar{n}_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \frac{X_{hi\ell k}}{P_{h\ell}} \right)_{hi} \\ & = \sum_{i=1}^{M_h} \rho_p(X_{hijk}, X_{hi\ell k})_{hi} \sigma_p^2(X)_{hi} \\ & E \left[ \frac{1}{n_{2h}-1} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \left( \frac{Y_{hij}}{P_{hj}} - \frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hij}}{P_{hj}} \right)^2 \right] = \sigma_p^2(Y)_h \end{aligned}$$

or

$$\begin{aligned} & E \left[ \sum_{i=1}^{m_{2h}} \frac{1}{\bar{n}_{2h}-1} \sum_{j=1}^{\bar{n}_{2h}} \left( \frac{Y_{hij}}{P_{hj}} - \frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hij}}{P_{hj}} \right) \right]_{hi} = \sum_{i=1}^{M_h} \sigma_p^2(Y)_h \\ & E \left[ \frac{1}{n_{2h}-1} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \left( \frac{Y_{hij}}{P_{hj}} - \frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hij}}{P_{hj}} \right) \right]_h \\ & \left( \frac{Y_{hi\ell}}{P_{hi}} - \frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{X_{hi\ell}}{P_{hi}} \right)_h \\ & = \rho_p(Y_{hij}, Y_{hi\ell})_h \sigma_p^2(Y)_h \end{aligned}$$

or

$$\begin{aligned} & E \left[ \sum_{i=1}^{m_{2h}} \frac{1}{\bar{n}_{2h}-1} \sum_{j=1}^{\bar{n}_{2h}} \left( \frac{Y_{hij}}{P_{hj}} - \frac{1}{\bar{n}_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hij}}{P_{hj}} \right) \right]_{hi} \\ & \left( \frac{Y_{hi\ell}}{P_{h\ell}} - \frac{1}{\bar{n}_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hi\ell}}{P_{h\ell}} \right)_{hi} \\ & = \sum_{i=1}^{M_h} \rho_p(Y_{hij}, Y_{hi\ell})_{hi} \sigma_p^2(Y)_{hi} \\ & E \left[ \frac{1}{n_{2h}-1} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \left( \frac{X_{hijk}}{P_{hj}} - \frac{1}{Kn_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} \right) \right]_h \\ & \times \left( \frac{Y_{hi\ell}}{P_{hi}} - \frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hi\ell}}{P_{hi}} \right)_h = \rho_p(Y_{hij}, Y_{hi\ell})_h \sigma_p^2(Y)_h \end{aligned}$$

or

$$\begin{aligned} & E \left[ \sum_{i=1}^{m_{2h}} \frac{1}{\bar{n}_{2h}-1} \sum_{j=1}^{\bar{n}_{2h}} \left( \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} - \frac{1}{K\bar{n}_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \frac{X_{hijk}}{P_{hj}} \right) \right]_{hi} \\ & \times \left( \frac{Y_{hij}}{P_{hj}} - \frac{1}{n_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Y_{hij}}{P_{hj}} \right)_{hi} = \sum_{i=1}^{M_h} \rho_p(X_{hijk}, Y_{hij})_{hi} \sigma_p(X)_{hi} \sigma_p(Y)_{hi} \end{aligned}$$

For unbiased estimation of

$$\sigma^2(\alpha)_h, \sigma_p^2(\beta)_h \text{ and } \sigma_p^2(d)_{hi}$$

we give the following definitions. The calculation of estimates of the simple response variances in Eq. (27) can be done by considering the following definitions on  $S(n_{2h})$ :

$$\begin{aligned} X_{hijk} &= Y_{hij} + \alpha_{hi} + \beta_{hj} + d_{hijk}; \\ h &= 1, 2, \dots, H; i = 1, 2, \dots, m_{2h} \end{aligned}$$

$$j = 1, 2, \dots, \bar{n}_{2h}; k = 1, 2, \dots, K \quad (28)$$

$$\bar{X}_{hij} = \frac{1}{K} \sum_{k=1}^K X_{hijk} = Y_{hij} + \alpha_{hi} + \beta_{hj} \quad (29)$$

$$\bar{X}_{hi\bullet} = \frac{1}{\bar{n}_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \bar{X}_{hij\bullet} = \bar{Y}_{hi\bullet} + \alpha_{hi} + \bar{\beta}_h \quad (30)$$

$$\bar{X}_{h\bullet} = \frac{1}{m_{2h}} \sum_{i=1}^{m_{2h}} \bar{X}_{hi\bullet} = \bar{Y}_{h\bullet} + \bar{\alpha}_h + \bar{\beta}_h \quad (31)$$

Hence, Eq. (28-29):

$$X_{hijk} - \bar{X}_{hij\bullet} = d_{hijk} \quad (32)$$

Eq. (29-30):

$$\bar{X}_{hij\bullet} - \bar{X}_{hi\bullet} = (Y_{hij} - \bar{Y}_{hi\bullet}) + (\beta_{hj} - \bar{\beta}_h)$$

$$\bar{x}_{hij\bullet} = y_{hij} + (\beta_{hj} - \bar{\beta}_h)$$

and hence

$$\beta_{hj} - \bar{\beta}_h = \bar{x}_{hij\bullet} - y_{hij} \quad (33)$$

Where:

$$\bar{x}_{hij\bullet} = \bar{X}_{hij\bullet} - \bar{X}_{hi\bullet} \text{ and } y_{hij} = Y_{hij} - \bar{Y}_{hi\bullet}$$

Let,

$$Z_{hij} = \bar{x}_{hij\bullet} - y_{hij} = \beta_{hj} - \bar{\beta}_h \text{ with } \bar{Z}_{hi\bullet} = 0$$

Then,

$$\sigma_p^2(\beta)_h = \text{Var} \left( \frac{Z_{hij}}{P_{hj}} \right)_{hi} = \sigma_p^2(Z)_{hi}$$

and hence

$$\begin{aligned} & E \left[ \frac{1}{n_{2h}-1} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \left( \frac{Z_{hij}}{P_{hj}} - \frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Z_{hij}}{P_{hj}} \right)^2 \right] \\ & = E \left[ \frac{1}{n_{2h}-1} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \left( \frac{Z_{hij}}{P_{hj}} \right)^2 \right]_{hi} = \sigma_p^2(Z)_h = \sigma_p^2(\beta)_h \end{aligned}$$

$$\begin{aligned} & E \left[ \frac{1}{n_{2h}-1} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \left( \frac{Z_{hij}}{P_{hj}} - \frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Z_{hij}}{P_{hj}} \right) \right]_h \\ & \times \left( \frac{Z_{hi\ell}}{P_{h\ell}} - \frac{1}{n_{2h}} \sum_{i=1}^{m_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \frac{Z_{hi\ell}}{P_{h\ell}} \right)_h = \rho_p(\beta_{hj}, \beta_{hi})_h \end{aligned}$$

Eq. (30-31):

$$\bar{X}_{hi\bullet} - \bar{X}_{hi\bullet\bullet} = (\bar{Y}_{hi\bullet} - \bar{Y}_{h\bullet}) + (\alpha_{hj} - \bar{\alpha}_{h\bullet})$$

or

$$\bar{x}_{hi} = \bar{y}_{hi} + (\alpha_{hj} - \bar{\alpha}_{h\bullet})$$

or

$$\alpha_{hi} - \bar{\alpha}_{h\bullet} = \bar{x}_{hi} - \bar{y}_{hi} \quad (34)$$

Where:

$$\bar{x}_{hi} = \bar{X}_{hi\bullet} - \bar{X}_{h\bullet}$$

and

$$\bar{y}_{hij} = \bar{Y}_{hi} - \bar{Y}_{h\bullet}$$

Let,

$$w_{hi} = \bar{x}_{hi} - \bar{y}_{hi} = \alpha_{hi} - \bar{\alpha}_{h\bullet} \text{ with } \bar{w}_{h\bullet} = 0$$

Then, since  $\sigma^2(\alpha)_h$  cannot be estimated unbiasedly, we find an unbiased estimator of

$$S^2(\alpha)_h = \frac{1}{M_h - 1} \sum_{i=1}^{M_h} (w_{hi} - \bar{w}_h)^2$$

by considering

$$E \left[ \frac{1}{m_{2h} - 1} \sum_{i=1}^{m_{2h}} (w_{hi} - \bar{w}_h)^2 \right] = \left[ \frac{1}{m_{2h} - 1} \sum_{i=1}^{m_{2h}} w_{hi}^2 \right]_{\bar{w}_h} = S^2(\alpha)_h$$

From Eq. (32)

$$E \left[ \frac{1}{K\bar{n}_{2h} - 1} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \left( \frac{d_{hijk}}{P_{hj}} - \frac{1}{K\bar{n}_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \frac{d_{hijk}}{P_{hj}} \right)^2 \right]_{hi} = \sigma_p^2(d)_{hi}$$

or

$$E \left[ \frac{1}{K\bar{n}_{2h} - 1} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \left( \frac{d_{hijk}}{P_{hj}} - \frac{1}{K\bar{n}_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \frac{d_{hijk}}{P_{hj}} \right) \times \left( \frac{d_{hijk}}{P_{hi}} - \frac{1}{K\bar{n}_{2h}} \sum_{j=1}^{\bar{n}_{2h}} \sum_{k=1}^K \frac{d_{hijk}}{P_{hi}} \right) \right] = \rho_p(d_{hijk}, d_{hi\neq k})_{hi} \sigma_p^2(d)_{hi}$$

## RESULTS AND DISCUSSION

We regret that it is difficult to find data from a survey based on the design discussed in this study for illustration of the various computations suggested by the previous results. As observed by Hansen *et al.* (1961), experiment that can generate the necessary data for the illustration are often highly expensive. The main aim of

this study, therefore, is to provide the mathematical framework within which statistical inferences can be made about the contributions of specified sources of survey error to respond variance in double sampling with unequal probabilities. To facilitate application of the method discussed in the study, we suggest that the survey data should be tabulated in a 4-dimensional table with h-k as the column headings.

For the purpose of choice between the estimators in Eq. 5 and  $\hat{B}$  in Eq. (8), we subtract  $v(\hat{B})$  in Eq. 27 from  $v(\hat{B}^*)$  in 6 and obtain the result

$$v(\hat{B}^*) - v(\hat{B}) = \sum_{h=1}^H \left( \frac{1}{n_{2h}} - \frac{1}{n_{1h}} \right) \sum_{i=1}^{M_h} \sigma_p^2(X)_{hi} \left[ 1 - 2\rho_p(X_{hij}, Y_{hij}) \frac{\sigma_p(Y)_{hi}}{\sigma_p(X)_{hi}} \right]$$

This result shows that  $v(\hat{B}^*) \leq v(\hat{B})$  if, in every interviewer assignment,

$$\rho_p(X_{hij}, Y_{hij}) \geq \frac{\sigma_p(X)_{hi}}{2\sigma_p(Y)_{hi}}$$

The implication of this is that if in every intervener assignment the biases are minimal so that  $X_{hij}$  has a high positive correlation with  $Y_{hij}$  i.e., both  $X_{hij}$  and  $Y_{hij}$  have the same distribution with  $\sigma_p(X) \approx \sigma_p(Y)$  the estimator  $\hat{B}^*$  based on information from the subsample only should be preferred to  $\hat{B}$  which is based on information from both samples. If as a result of non-stationary distribution of indiscriminate biases in  $X$ ,  $\sigma_p(X)_{hi} \gg \sigma_p(Y)$ , the observed value of  $\rho_p(X_{hij}, Y_{hij})$  may be

$$< \frac{\sigma_p(X)_{hi}}{2\sigma_p(Y)_{hi}}$$

within the interviewer assignments. In such a situation which may be common to most practical social survey situations,  $\hat{B}$  in Eq. (8) is preferable to  $\hat{B}^*$  in Eq. (5) above. The choice of  $\hat{B}$  in this case enables the investigator to include in his estimation model more of the highly variable measurements  $\{X_{hijk}\}$  than the precise and less variable measurements  $\{Y_{hij}\}$ . This agrees with the precision rule is stratified random sampling with proportional allocation.

## CONCLUSION

Equations 19 and 27 show that under the sample design used in this study, the contribution from correlated



factors to the total response variance is also substantial as observed by Hansen *et al.* (1961) under a different sample design. Repetition of the survey for  $K > 1$  times is necessary for the calculation of the error terms. However, the magnitude of  $K$  affects the contribution from correlated factors positively more than it affects the contribution from each of the other sources of variation. For this reason,  $K$  can be limited to 2. Since, the ratio of  $\sigma_p(X)$  to  $\sigma_p(Y)$  suggests the amount of distortions that make the distribution of  $X$  different from that of  $Y$ , which is desired, the ratio  $\sigma_p(X)/\sigma_p(Y)$  can be calculated from a pilot survey data as a basis of choice between  $\hat{p}^*$  in Eq. 5 and  $\hat{p}$  Eq. in 8 above.

### REFERENCES

- Fellegi, J.P., 1964. Response variation and its estimation. *J. Am. Stat. Assoc.*, 59: 1016-1041.
- Hansen, M.H., W.N. Hurwitz and M.A. Bershad, 1961. Measurement error in censuses and surveys. *Bull. Int. Stat. Inst.*, 38: 359-374.
- Hansen, M.H., W.N. Hurwitz and W.G. Madow, 1953. *Sampling Survey Methods and Theory*, Vol. II, John Wiley and Sons Inc; New York. ISBN: 47134947X.
- Hansen, M.H., W.N. Hurwitz, E.S. Marks and W.F. Mauldin, 1951. Response errors in surveys. *J. Am. Stat. Assoc.*, 46: 147-190.
- Kish, L., 1965. *Survey Sampling*. John Wiley and Sons Inc., New York. ISBN: 047148900X.
- Koch, G., 1973. An alternative approach to multivariate response error models for sample survey data with application to estimators involving subclass means. *J. Am. Stat. Assoc.*, 68: 906-913.
- Mahalanobis, P.C., 1946. Recent experiments in statistical sampling in the Indian Statistical Institute. *J. R. Stat. Soc.*, 109: 325-380.
- Raj, D., 1956. Some estimators in sampling with varying probabilities without replacement. *J. Indian Soc. Agric. Stats.*, 6: 127-133.
- Raj, D., 1965. Response bias in farmers' reports. *Sankhyā B*, pp: 279-282.
- Raj, D., 1968. *Sampling Theory*. McGraw-Hill, New York.
- Talukder, M.A.H., 1975. Response bias and difference estimate in double sampling. *Metrika*, 22: 65-75.
- Udofia, G.A., 2006. Response variance in double sampling for inclusion probabilities. *Int. J. Natl. Applied Sci.*, 1 (1): 89-95.
- Udofia, G.A., 2008. Estimation of response bias based on information from both samples in double sampling for inclusion probabilities. *J. Math. Sci., International Centre for Advance Studies, India* 19 (2): 127-137.
- Warwick, D.P. and C.P. Lininger, 1975. *The Sample Survey: Theory and Practice*. McGraw-Hill Book Company, New York. ISBN: 0-07-068396-4, 0-07-068395-6 (pbk).