Estimation in Elusive Populations Using Multiple Frames and Two-Phase Multiple Frames in the Presence of Measurement and Response Errors

Mutanu Beth[1], Kahiri James[2], Odongo Leo[2]

[1]Mathematics and Actuarial Science Department, South Eastern Kenya University, Kitui, Kenya

[2]Mathematics and Actuarial Science Department, Kenyatta University, Nairobi, Kenya

Email address:
mutanubeth24@gmail.com (M. Beth), jmkahiri2012@gmail.com (K. James), odongo.leo@ku.ac.ke (O. Leo)

Abstract: Accurate survey data is important for planning and decision making. The presence of measurement and response errors in surveys has been known to negatively affect the efficiency of estimates as well as to create biases in estimates. It is important to investigate the effects of measurement and response errors when computing survey data so as to obtain reliable information for use by statisticians and policy makers. Unavailability of a sampling frame in a survey for elusive populations has led to the application of multiple frames in sample selection processes. This paper investigates the effect of measurement and response errors in population estimation under multiple and two-phase multiple frames for elusive populations. The effect of random errors and biases from systematic errors on simple and correlated response variances under various levels of multiplicity adjustment factor in multiple frames is carried out. A numerical example is given assuming simple random sampling. The net effect of the errors has been found to inflate simple and correlated response variances and hence overestimation of the variances under different variance estimators. It is therefore recommended that both measurement and response errors be put into consideration when designing and carrying out a survey for more accurate results.

Keywords: Elusive Populations, Multiple Frames, Measurement Errors, Multiple Adjustment Factor, Simple Response Variance, Correlated Response Variance

1. Introduction

When only a part of the population is studied to make inferences about a target population, sampling errors may occur. These errors would therefore not exist if a census is conducted. In the course of collecting data, errors may arise through the measurement process, interviewing or during observations. The respondent may fail to respond leading to non-sampling errors associated with the study variable.

Estimation in elusive population in which a direct sampling frame may not exist can be done by use of multiple frames (MF) surveys [1]. Multiple frames are useful for reducing costs for a given precision constraint and improving coverage, under-or over-coverage [2]. To overcome this problem, the original frame is supplemented with additional frames so that the union of all frames covers the target population. In this case some units may be present in more than one frame, leading to over-coverage. Duplication in the same frame may also occur leading to over-coverage.

Thus imperfect frames can be dealt with by use of MF surveys where two or more listings of units is simultaneously used for sample selection [3]. For instance, we can have a three-frame sampling design by carrying out a MF survey of homeless population where eligibility is from homeless shelters, soup kitchens and street locations [4].

Thus MF set up correspond to sampling individually at locations and has been applied in European migration studies [5]. MF surveys allows use of more flexible strategy than conventional single frame survey as well as use of different modes of data collection. For example, in one frame one can use face-to-face interviews and in another frame use email-questionnaires with an aim of controlling coverage survey costs,

response rates and estimation accuracy. Singh and Mecatti [6, 7] provided a simplified and unified review of available MF estimation methods by using the multiplicity approach to MF estimation when there are no measurement and response errors.

2. Estimation in Multiple Frames When There Are Measurement and Response Errors

For each frame, it is possible to partition the list of units into disjoint sets determined by the membership of the units in other frames. These disjoint sets are called domains of the frame. For instance, if there are two frames say A and B the disjoint sets or the domains of A are two, those in A and those in both A and B i.e. a (A) and a (AB). Similarly, B has two domains b (B) and b (AB) where a (AB) and b (AB) are same set.

In general, if there are say L frames, then the number of domains in a frame will have;

$$\binom{L-1}{0}+\binom{L-1}{1}+........+\binom{L-1}{L-1}=\sum_{r=0}^{L-1}\binom{L-1}{r}=(1+1)^{L-1}=2^{L-1} \text{ domains.}$$

The number of domains in a frame is equal to 2^{L-1} and the total number of domains in all the frames is equal to 2^L-1. In MF surveys we use the frame membership indicator $I_{k\in l}, l=1,...,L$ which is equal to 1 if population unit k is included in frame l and 0 otherwise. That is,

$$I_{lk}=\begin{cases}1 & \text{if } k^{th} \text{ population unit belongs to frame } l \\ 0, & \text{otherwise}\end{cases}$$

Let m_d be the number of frames a unit in domain d belongs to. Let the size of frame l be $N_l=\sum_{k=1}^{L}I_{k\in l}$. Note that $\sum_{l=1}^{L}N_l\geq N$.

The population mean can therefore be written as,

$$\bar{\mu}=\frac{1}{N}\sum_{l=1}^{L}\frac{1}{m_d}\sum_{k\in s_{ld}}\mu_{lk} \tag{1}$$

Now consider that samples are taken one from each frame and denote the samples as S_l from the lth frame. The sample membership can be partitioned into domain membership within frame l as S_{ld}. The estimator is therefore,

$$\hat{\bar{\mu}}=\frac{1}{N}\sum_{l=1}^{L}\sum_{k\in s_{ld}}\frac{\mu_{lk}}{m_d\pi_{lk}}$$

Where π_{lk} is the inclusion probability of unit k in the sample.

We consider estimation in multiple frames when there are measurement and response errors. In the estimation process the observed value for the kth unit in frame l can be modeled as $y_{lk}=\mu_{lk}+e_{lk}$

where the observation made is y_{lk} instead of μ_{lk}. The estimator of the population mean is therefore

$$\hat{\bar{y}}=\frac{1}{N}\sum_{l}^{L}\sum_{k\in d}\frac{y_{lk}}{m_d\pi_{lk}}$$

$$E\left(\hat{\bar{y}}\right)=E_{pm}\left(\hat{\bar{y}}\right)=E_p\frac{1}{N}\sum_{l=1}^{L}\sum_{k\in s_{ld}}\frac{E_m(y_{lk})}{m_d\pi_{lk}}$$

$$=\frac{1}{N}\sum_{l=1}^{L}E_p\sum_{k\in s_{ld}}\frac{\mu_{lk}^*}{m_d\pi_{lk}} \text{ where } E_m(y_{lk})=\mu_{lk}+e_{lk}=\mu_{lk}^*$$

$$E\left(\hat{\bar{y}}\right)=\frac{1}{N}\sum_{l=1}^{L}\sum_{k\in s_{ld}}\frac{\mu_{lk}^*}{m_d\pi_{lk}}E\left(I_{lk}\right)$$

$$=\frac{1}{N}\sum_{l=1}^{L}\sum_{k\in s_{ld}}\frac{\mu_{lk}^*}{m_k}$$

The variance of $\hat{\bar{y}}$ is

$$v\left(\hat{\bar{y}}\right)=V_{pm}\left(\hat{\bar{y}}\right)=V_pE_m\left(\hat{\bar{y}}\right)+E_pV_m\left(\hat{\bar{y}}\right)$$

where

$$V_pE_m\left(\frac{1}{N}\sum_{l=1}^{L}\sum_{k\in s_{ld}}\frac{y_{lk}}{m_d\pi_{lk}}\right)$$

$$=V_p\left(\frac{1}{N}\sum_{l=1}^{L}\sum_{k\in s_{ld}}\frac{E_m(y_{lk})}{m_d\pi_{lk}}\right)$$

$$=V_p\left(\frac{1}{N}\sum_{l=1}^{L}\sum_{k\in s_{ld}}\frac{\mu_{lk}^*I_{lk}}{m_d\pi_{lk}}\right)$$

$$=\frac{1}{N^2}\left\{\sum_{l=1}^{L}\frac{1}{m_d^2}\left[\sum_{k=1}^{N_l}\frac{\mu_{lk}^{*2}}{\pi_{lk}^2}(\pi_{lk}-\pi_{lk}^2)+\sum_{k\neq k'}\sum\frac{\mu_{lk}^*\mu_{lk'}^*}{\pi_{lk}\pi_{lk'}}(\pi_{lkk'}-\pi_{lk}\pi_{lk'})\right]\right. \\ \left.+\sum_{l\neq l'}\sum\frac{1}{m_dm_{d'}}\sum_{k\in s_{ld}}\sum_{k'\in s_{l'd'}}\frac{\mu_{lk}^*\mu_{l'k'}^*}{\pi_{lk}\pi_{l'k'}}(\pi_{ll'kk'}-\pi_{lk}\pi_{l'k'})\right\} \tag{2}$$

and

$$E_pV_m\left(\hat{\bar{y}}\right)$$

$$V_m\left(\hat{\bar{y}}\right)=V_m\left(\frac{1}{N}\sum_{l=1}^{L}\sum_{k\in s_{ld}}\frac{y_{lk}}{m_d\pi_{lk}}\right)$$

$$=\frac{1}{N^2}\left[\sum_{l=1}^{L}\frac{1}{m_d^2}\left(\sum_{k\in s_{ld}}\frac{V(e_{lk})}{\pi_{lk}}+\sum_{k\neq k'}\sum\frac{COV(e_{lk},e_{lk'})}{\pi_{lk}\pi_{lk'}}\right)\right]$$

$$=\frac{1}{N^2}\sum_{l=1}^{L}\frac{1}{m_d^2}\left[\sum_{k=1}^{N_l}\frac{\sigma_{lk}^2}{\pi_{lk}^2}+\sum_{k\neq k'}\sum\frac{\sigma_{lkk'}}{\pi_{lk}\pi_{lk'}}\right]$$

and

$$E_p V_m \left(\hat{\bar{y}} \right) = E_p \left\{ \frac{1}{N^2} \sum_{l=1}^{L} \frac{1}{m_d^2} \left[\sum_{k=1}^{N_l} \frac{\sigma_{lk}^2}{\pi_{lk}^2} I_k + \sum_{k \neq k'} \sum \frac{\sigma_{lkk'}}{\pi_{lk} \pi_{lk'}} I_{lk} I_{lk'} \right] \right\}$$

$$= \frac{1}{N^2} \sum_{l=1}^{L} \frac{1}{m_d^2} \left[\sum_{k=1}^{N_l} \frac{\sigma_{kl}^2}{\pi_{lk}^2} \pi_{lk} + \sum_{k \neq k'} \sum \frac{\sigma_{lkk'}}{\pi_{lk} \pi_{lk'}} \pi_{lkk'} \right]$$

$$= \frac{1}{N^2} \sum_{l=1}^{L} \frac{1}{m_d^2} \left[\sum_{k=1}^{N_l} \frac{\sigma_{lk}^2}{\pi_{lk}} + \sum_{k \neq k'} \sum \frac{\sigma_{lkk'}}{\pi_{lk} \pi_{lk'}} \pi_{lkk'} \right]$$

3. Generalized-multiplicity Adjusted Horvitz-thompson Estimator in Presence of Measurement and Response Errors

The generalized-multiplicity adjusted Horvitz-Thompson estimator was introduced by Singh and Mecatti (2009, 2011). This is defined as

$$\hat{y}_{GMHT} = \sum_{l=1}^{L} \sum_{k \in s_{ld}} \frac{y_{lk}}{\pi_{lk}} \beta_{lk} \tag{3}$$

where β_{kl} ensures that y_k is counted once despite the number of replications. The multiplicity adjustment factor β_{kl} is such that $\Sigma \beta_{kl} = 1$

If $\beta_{kl} = m_d^{-1}$, then $\hat{y}_{GMHT} = \sum_{l=1}^{L} \sum_{k \in s_{ld}} \frac{y_{lk}}{m_d \pi_{lk}}$ as observed earlier.

For dual frames we saw that each frame has two domains.

$$\hat{y}_{GMHT} = \sum_{k \in SA} \frac{y_k}{\pi_{kA}} \beta_k I_{k \in a(A)} + \sum_{k \in SA} \frac{y_k}{\pi_{kA}} \beta_k I_{k \in a(AB)} + \sum_{k \in SB} \frac{y_k}{\pi_{kB}} \beta_k I_{k \in b(B)} + \sum_{k \in SB} \frac{y_k}{\pi_{kB}} \beta_k I_{k \in b(AB)}$$

If $\beta_{kl} = m_k^{-1}$, then those in a (A) and b (B) have $m_k = 1$ and those in a (AB) and b (AB) have $m_k = 2$

This gives $\hat{y}_{GMHT} = \sum_{k \in SA} \frac{y_k}{\pi_{kA}} I_{k \in a(A)} + \sum_{k \in SA} \frac{y_k}{2\pi_{kA}} I_{k \in a(AB)} + \sum_{k \in SB} \frac{y_k}{\pi_{kB}} I_{k \in b(B)} + \sum_{k \in SB} \frac{y_k}{2\pi_{kB}} I_{k \in b(AB)}$

Using the Kalton Anderson multiplicity adjustment factor we get [8],

$$\hat{y}_{kA} = \sum_{k \in SA} \frac{y_k}{\pi_A} I(a(A)) + \sum_{k \in SA} \frac{y_k}{\pi_A} \left(\frac{\pi_A}{\pi_A + \pi_B} \right) I(a(AB)) + \sum_{k \in SB} \frac{y_k}{\pi_B} \left(\frac{\pi_B}{\pi_A + \pi_B} \right) I(b(AB)) + \sum_{k \in SB} \frac{y_k}{\pi_B} I(b(B))$$

$$= \hat{y}_{a(A)} + \frac{\pi_A}{\pi_A + \pi_B} \hat{y}_{a(AB)} + \frac{\pi_B}{\pi_A + \pi_B} \hat{y}_{b(AB)} + \hat{y}_{b(B)} \tag{4}$$

where

$$\hat{y}_{a(A)} = \sum_{k \in SA} \frac{y_k}{\pi_{k(A)}} I_{k \in a(A)}$$

$$\hat{y}_{a(AB)} = \sum_{k \in SA} \frac{y_k}{\pi_{k(A)}} I_{k \in a(AB)}$$

$$\hat{y}_{b(B)} = \sum_{k \in SB} \frac{y_k}{\pi_{k(B)}} I_{k \in b(B)}$$

$$\hat{y}_{b(AB)} = \sum_{k \in SB} \frac{y_k}{\pi_{k(B)}} I_{k \in b(AB)}$$

$$\hat{y}_{kA} = \sum_{k \in SA} \frac{y_k}{\pi_k} I_{k \in a(A)} + \frac{\pi_A}{\pi_A + \pi_B} \sum_{k \in SA} \frac{y_k}{\pi_k} I_{k \in a(AB)} + \frac{\pi_B}{\pi_A + \pi_B} \sum_{k \in SB} \frac{y_k}{\pi_k} I_{k \in b(AB)} + \sum_{k \in SB} \frac{y_{lk}}{\pi_{lk}} I_{k \in b(B)}$$

$$= \sum_{k \in SA} \frac{y_k}{\pi_k} \left[I_{k \in a(A)} + \frac{\pi_A}{\pi_A + \pi_B} I_{k \in a(AB)} \right] + \sum_{k \in SB} \frac{y_k}{\pi_k} \left[I_{k \in b(B)} + \frac{\pi_B}{\pi_A + \pi_B} I_{k \in b(AB)} \right]$$

$$(5)$$

This can be extended to three frames as shown below

$$\hat{y}_{kA3} = \hat{y}_{a(A)} + \frac{\pi_A}{\pi_A + \pi_B} \hat{y}_{a(AB)} + \frac{\pi_B}{\pi_A + \pi_B} \hat{y}_{b(AB)} + \frac{\pi_A}{\pi_A + \pi_C} \hat{y}_{a(AC)} + \hat{y}_{b(B)} + \frac{\pi_B}{\pi_B + \pi_C} \hat{y}_{b(BC)} + \frac{\pi_C}{\pi_B + \pi_C} \hat{y}_{c(BC)}$$

$$+ \frac{\pi_C}{\pi_A + \pi_C} \hat{y}_{c(AC)} + \hat{y}_{c(C)} + \frac{\pi_A}{\pi_A + \pi_B + \pi_C} \hat{y}_{a(ABC)} + \frac{\pi_B}{\pi_A + \pi_B + \pi_C} \hat{y}_{b(ABC)} + \frac{\pi_C}{\pi_A + \pi_B + \pi_C} \hat{y}_{c(ABC)}$$

this can be written as

$$\hat{y}_{kA3} = \sum_{k \in SA} \frac{y_k}{\pi_k} \left[I_{k \in a(A)} + \frac{\pi_A}{\pi_A + \pi_B} I_{k \in a(AB)} + \frac{\pi_A}{\pi_A + \pi_C} I_{k \in a(AC)} + \frac{\pi_A}{\pi_A + \pi_B + \pi_C} I_{k \in a(ABC)} \right]$$

$$+ \sum_{k \in SB} \frac{y_k}{\pi_k} \left[I_{k \in b(B)} + \frac{\pi_B}{\pi_A + \pi_B} I_{k \in b(AB)} + \frac{\pi_B}{\pi_B + \pi_C} I_{k \in b(BC)} + \frac{\pi_B}{\pi_A + \pi_B + \pi_C} I_{k \in b(ABC)} \right]$$

$$(6)$$

$$+ \sum_{k \in SC} \frac{y_k}{\pi_k} \left[I_{k \in c(C)} + \frac{\pi_C}{\pi_A + \pi_C} I_{k \in c(AC)} + \frac{\pi_C}{\pi_B + \pi_C} I_{k \in c(BC)} + \frac{\pi_C}{\pi_A + \pi_B + \pi_C} I_{k \in c(ABC)} \right]$$

A generalization to L frames would be

$$\hat{y}_{kA} = \sum_{l=1}^{L} \sum_{k \in S_L} \frac{y_{lk}}{\pi_{lk}} \left\{ \sum_{d=1}^{D} \left(\frac{\pi_{kL} I_{k \in d}}{\sum_{l=1}^{L} \pi_{kL} I_{k \in d}} \right) \right\}$$

$$(7)$$

This implies the multiplicity adjustment factor is

$$\beta_{lk}^{kA} = \frac{\pi_{lk} I_{k \in d}}{\sum_{d=1}^{D} \sum_{l=1}^{L} \pi_{lk} I_{k \in d}}$$

$$= \pi_{lk} \sum_{d=1}^{D} \frac{I_{k \in d}}{\sum_{l=1}^{L} \pi_{lk} I_{k \in d}}$$

In the presence of measurement errors expectation of the estimator is

$$
\begin{aligned}
E_{pm}\left(\hat{y}_{GMHT}\right) &= E_{pm}\left(\sum_{l=1}^{L}\sum_{k\in S_l}\frac{y_{lk}}{\pi_{kl}}\beta_{lk}\right)\\
&= E_p\left[\sum_{l=1}^{L}\sum_{k\in S_l}\frac{\mu_{LK}^*}{\pi_{lk}}\beta_{lk}\right] \qquad \text{where } \mu_{LK}^* = E_m\left(y_{lk}\right)\\
&= \sum_{l=1}^{L}\sum_{k=1}^{N_L}\frac{\mu_{lk}^*\beta_{lk}}{\pi_{lk}}E_p\left(I_{k\in S_l}\right)\\
&= \sum_{l=1}^{L}\sum_{k=1}^{N_L}\frac{\mu_{lk}^*\beta_{lk}}{\pi_{lk}}\pi_{lk} = \sum_{l=1}^{L}\sum_{k=1}^{N_L}\mu_{lk}^*\beta_{lk}
\end{aligned}
\tag{8}
$$

The variance is given by

$$
\begin{aligned}
V_{pm}\left(\hat{y}_{GMHT}\right) &= V_p E_m\left(\hat{y}_{GMHT}\right) + E_p V_m\left(\hat{y}_{GMHT}\right)\\
V_p E_m\left(\hat{y}_{GMHT}\right) &= V_p\left(\sum_{l=1}^{L}\sum_{k\in S_l}\frac{\mu_{lk}^*\beta_{lk}}{\pi_{lk}}\right)\\
&= V_p\left(\sum_{l=1}^{L}\sum_{k=1}^{N_L}\frac{\mu_{lk}^*}{\pi_{lk}}\beta_{lk}I_{lk}\right)\\
&= \sum_{l=1}^{L}\left[\sum_{k=1}^{N_L}\frac{\mu_{lk}^{*\,2}}{\pi_{lk}^2}\beta_{lk}^2\pi_{lk}\left(1-\pi_{lk}\right) + \sum_{k\neq k'}\sum\frac{\mu_{lk}^*\mu_{lk'}^*\beta_{lk}\beta_{lk'}}{\pi_{lk}\pi_{lk'}}\left(\pi_{lkk'}-\pi_{lk}\pi_{lk'}\right)\right]\\
&\quad + \sum_{l\neq l'}\sum\sum_{k=1}^{N_L}\sum_{k'=1}^{N_{L'}}\frac{\mu_{lk}^*\mu_{l'k'}^*}{\pi_{lk}\pi_{l'k'}}\left(\pi_{ll'kk'}-\pi_{lk}\pi_{l'k'}\right)\beta_{lk}\beta_{l'k'}
\end{aligned}
\tag{9}
$$

This is sampling variance assuming the general case where there could be correlation between frames.

$$
\begin{aligned}
E_p V_m\left(\hat{y}_{GMHT}\right) &= E_p\left[\sum_{l=1}^{L}\left(\sum_{k\in S_l}\frac{\sigma_{lk}^2\beta_{lk}^2}{\pi_{lk}^2} + \sum_{l\neq l'}\sum_{k\in S_l}\sum\frac{\pi_{lk'}}{\pi_{lk}\pi_{lk'}}\beta_{lk}\beta_{lk'}\right)\right] + E_p\left[\sum_{l\neq l'}\sum\sum_{k\in S_l}\sum_{k\in S_{l'}}\frac{\sigma_{ll'kk'}}{\pi_{lk}\pi_{l'k'}}\beta_{lk}\beta_{l'k'}\right]\\
&= \sum_{l=1}^{L}\left[\sum_{k=1}^{N_L}\frac{\sigma_{lk}^2\beta_{lk}^2}{\pi_{lk}} + \sum_{k\neq k'}\sum\frac{\sigma_{lkk'}}{\pi_{lk}\pi_{lk'}}\beta_{lk}\beta_{lk'}\pi_{lkk'}\right] + \sum_{l\neq l'}\sum\sum_{k=1}^{N_L}\sum_{k'=1}^{N_{L'}}\frac{\pi_{ll'kk'}}{\pi_{lk}\pi_{l'k'}}\sigma_{ll'kk'}\beta_{lk}\beta_{l'k'}
\end{aligned}
\tag{10}
$$

$$
V\left(\hat{y}_{GHMT}\right) = \sum_{l=1}^{L}\left[\sum_{k=1}^{N_L}\frac{\sigma_{lk}^2}{\pi_{lk}^2}\pi_{lk}\left(1-\pi_{lk}\right)\left(\pi_{lk}\sum_{d=1}^{D}\frac{I_{k\in d}}{\sum_{l=1}^{L}\pi_{lk}I_{k\in d}}\right)^2 + \sum_{k\neq k'}\sum\frac{\sigma_{lk}\sigma_{lk'}}{\pi_{lk}\pi_{lk'}}\left(\pi_{lk}-\pi_{lk}\pi_{lk'}\right)+\right.
$$

$$
\left.\sum_{l\neq l'}\sum\sum_{k=1}^{N_L}\sum_{k'=1}^{N_{L'}}\frac{\sigma_{lk}\sigma_{l'k'}}{\pi_{lk}\pi_{l'k}}\left(\pi_{ll'kk'}-\pi_{lk}\pi_{l'k'}\right)\left[\pi_{lk}\left(\sum_{d=1}^{D}\frac{I_{k\in d}}{\sum_{l=1}^{L}\pi_{kl}I_{k\in d}}\right)\right]\left[\pi_{l'k'}\left(\sum_{d=1}^{D}\frac{I_{k'\in d}}{\sum_{l=1}^{L}\pi_{k'l'}I_{k'\in d}}\right)\right]\right]
$$

$$
+ \sum_{l=1}^{L}\left[\sum_{k=1}^{N_L}\frac{\sigma_{lk}^2}{\pi_{lk}}\left(\pi_{lk}\sum_{d=1}^{D}\frac{I_{k\in d}}{\sum_{l=1}^{L}\pi_{lk}I_{k\in d}}\right)^2 + \sum_{l\neq l'}\sum\frac{\sigma_{lk}\sigma_{lk'}}{\pi_{lk}\pi_{lk'}}\pi_{lkk'} + \sum_{l\neq l'}\sum\sum_{k\neq k'}\sum\frac{\sigma_{ll'kk'}}{\pi_{lk}\pi_{l'k}}\left(\pi_{lk}\left(\sum_{d=1}^{D}\frac{I_{k\in d}}{\sum_{l=1}^{L}\pi_{lk}I_{k\in d}}\right)\right)\left(\pi_{l'k'}\left(\sum_{d=1}^{D}\frac{I_{k'\in d}}{\sum_{l=1}^{L}\pi_{l'k'}I_{k'\in d}}\right)\right)\right] \tag{11}
$$

4. Simple Random Sampling with Equal Sample Sizes Per Frame

With equal size simple random samples per frame, we get

$$V_{pm}\left(\hat{y}_{GMHT}\right)=V_pE_m\left(\hat{y}_{GMHT}\right)+E_pV_m\left(\hat{y}_{GMHT}\right)$$

where

$$V_pE_m\left(\hat{y}_{GMHT}\right)=\sum_{l=1}^{L}\left[\sum_{k=1}^{N_L}\frac{\mu_{lk}^*}{\left(\dfrac{n}{N_l}\right)^2}\beta_{lk}^2\frac{n}{N_l}\left(1-\frac{n}{N_l}\right)\;\;\;\;+\sum_{k\neq k'}\sum\mu_{lk}^*\frac{\mu_{lk'}^*\beta_{lk}\beta_{lk'}}{\dfrac{n}{N_l}\cdot\dfrac{n}{N_l}}\left(\frac{n}{N_l}\frac{(n-1)}{(N_l-1)}-\frac{n}{N_l}\cdot\frac{n}{N_l}\right)\right]$$

$$+\sum_{l\neq l'}\sum\sum_{k=1}^{N_L}\sum_{k'=1}^{N_{L'}}\frac{\mu_{lk}^*\mu_{l'k'}^*}{\dfrac{n}{N_l}\dfrac{n}{N_l}}\left(\frac{n}{N_l}\frac{(n-1)}{(N_l-1)}-\frac{n}{N_l}\cdot\frac{n}{N_l}\right)\beta_{lk}\beta_{l'K'}$$

$$=\sum_{l=1}^{L}\left[\frac{n(N_l-n)}{N^2}\sum_{k=1}^{N_L}\mu_{lk}^{*\,2}+\frac{N_l^2(n-N_l)}{N^2n(N_l-1)}\sum_{k\neq k'}\sum\mu_{lk}^*\mu_{lk'}^*\right]+\sum_{l=1}^{L}\frac{N_l^2(n-N_l)}{N^2n(N_l-1)}\sum_{k=1}^{N_L}\sum_{k'=1}^{N_{L'}}\mu_{lk}^*\mu_{l'k'}^* \tag{12}$$

$$E_pV_m\left(\hat{y}_{GMHT}\right)=\sum_{l=1}^{L}\left[\sum_{k=1}^{N_L}\frac{\sigma_{lk}^2\beta_{lk}^2}{\dfrac{n}{N_l}}+\sum_{k\neq k'}\sum_{}^{N_L}\frac{\sigma_{l'kk'}}{\dfrac{n}{N_l}\cdot\dfrac{n}{N_l}}\beta_{lk}\alpha_{lk}\frac{n(n-1)}{N_l(N_l-1)}\right]+\sum_{l\neq l'}\sum\sum_{k=1}^{N_L}\sum_{k'=1}^{N_{L'}}\frac{\dfrac{n(n-1)}{N_l(N_l-1)}}{\dfrac{n}{N_l}\cdot\dfrac{n}{N_l}}\sigma_{ll'kk}\beta_{lk}\beta_{l'k'}$$

$$=\sum_{l=1}^{L}\left[\frac{N_l^3}{nN^3}\sum_{k=1}^{N_L}\sigma_{lk}^2+\frac{N_l^3(n-1)}{N^3n(N_{l-1})}\sum\sum_{k\neq k'}^{N_L}\sigma_{lkk'}\right]+\sum_{l=1}^{L}\frac{N_l^3(n-1)}{N^3n(N_l-1)}\sum_{k=1}^{N_L}\sum_{k'=1}^{N_{L'}}\sigma_{ll'kk'} \tag{13}$$

Suppose the sample sizes are different for different frames. Then

$$V_{pm}\left(\hat{y}_{GMHT}\right)=V_pE_m\left(\hat{y}_{GMHT}\right)+E_pV_m\left(\hat{y}_{GMHT}\right)$$

where

$$V_pE_m\left(\hat{y}_{GMHT}\right)=\sum_{l=1}^{L}\left[\sum_{k=1}^{N_L}\frac{\mu_{lk}^{*\,2}}{\left(\dfrac{n_l}{N_l}\right)^2}\beta_{lk}^2\frac{n_l}{N_l}\left(1-\frac{n_l}{N_l}\right)+\sum_{k\neq k'}\sum\mu_{lk}^*\frac{\mu_{lk'}^*\beta_{lk}\beta_{lk'}}{\dfrac{n_l}{N_l}\cdot\dfrac{n_l}{N_l}}\left(\frac{n_l}{N_l}\frac{(n_l-1)}{(N_l-1)}-\frac{n_l}{N_l}\cdot\frac{n_l}{N_l}\right)\right]$$

$$+\sum_{l\neq l'}\sum\sum_{k=1}^{N_L}\sum_{k'=1}^{N_{L'}}\frac{\mu_{lk}^*\mu_{l'k'}^*}{\dfrac{n_l}{N_l}\dfrac{n_l}{N_l}}\left(\frac{n_l}{N_l}\frac{(n_l-1)}{(N_l-1)}-\frac{n_l}{N_l}\cdot\frac{n_l}{N_l}\right)\beta_{lk}\beta_{l'k'} \tag{14}$$

$$=\sum_{l=1}^{L}\left[\frac{n_l(N_l-n_l)}{N^2}\sum_{k=1}^{N_L}\mu_{lk}^{*\,2}+\frac{N_l^2(n_l-N_l)}{N^2n_l(N_l-1)}\sum_{k\neq k'}\sum\mu_{lk}^*\mu_{lk'}^*\right]+\sum_{l=1}^{L}\frac{N_lN_{l'}(n_l-N_l)}{N^2n_l(N_l-1)}\sum_{k=1}^{N_L}\sum_{k'=1}^{N_{L'}}\mu_{lk}^*\mu_{l'k'}^*$$

and

$$E_pV_m\left(\hat{y}_{GMHT}\right)=\sum_{l=1}^{L}\left[\sum_{k=1}^{N_L}\frac{\sigma_{lk}^2\beta_{lk}^2}{\dfrac{n_l}{N_l}}+\sum_{k\neq k'}\sum_{}^{N_L}\frac{\sigma_{lkk'}}{\dfrac{n_l}{N_l}\cdot\dfrac{n_l}{N_l}}\beta_{lk}\beta_{lk'}\frac{n_l(n_l-1)}{N_l(N_l-1)}\right]+\sum_{l\neq l'}\sum\sum_{k=1}^{N_L}\sum_{k'=1}^{N_{L'}}\frac{\dfrac{n_l(n_l-1)}{N_l(N_l-1)}}{\dfrac{n_l}{N_l}\cdot\dfrac{n_l}{N_l}}\sigma_{ll'kk}\beta_{lk}\alpha_{l'k'}$$

$$=\sum_{l=1}^{L}\left[\frac{N_l^3}{N^2n_l}\sum_{k=1}^{N_L}\sigma_{lk}^2+\frac{N_l^3(n_l-1)}{N^2n_l(N_l-1)}\sum\sum_{k\neq k'}^{N_L}\sigma_{lkk}\right]+\sum_{l=1}^{L}\frac{N_l^2N_{l'}(n_l-1)}{N^2n_l(N_l-1)}\sum_{k=1}^{N_L}\sum_{k'=1}^{N_{L'}}\sigma_{ll'kk'} \tag{15}$$

5. Estimation Under Two Phase Multiple Frames (MF) for Elusive Population When There Are Measurement and Response Errors

Now consider the l frames.

Assume that an original sample s_l' is selected from frame l where the probability of inclusion is say π_{lk1} and the joint inclusion probability is $\pi_{lkk'1}$

In phase two a subsample s_l is selected from s_l' with inclusion probability π_{lk2} and joint inclusion probability $\pi_{lkk'2}$. The overall inclusion probability is

$$\pi_{lk} = P(\text{k selected at phase 1})\,P(\text{k is selected at phase 2}\,/\,k\text{ selected at phase 1})$$
$$= \pi_{lk1}\pi_{lk2}$$

Using the generalized-multiplicity adjusted Horvitz-Thompson class of estimators.

$$\hat{Y}_{GMHT} = \sum_{l=1}^{L}\sum_{k\in s_l}\frac{y_{lk}}{\pi_{lk}}\beta_{lk}$$

We get $E\left(\hat{Y}_{GMHT}\right) = Ep_1 p_2 m\left(\hat{Y}_{GMHT}\right) = Ep_1 p_2 \left(\sum_{L}\sum_{k\in s_l}\frac{\mu_{lk}^{*}\beta_{lk}}{\pi_{lk}}\right)$

Where $E_m\left(y_{lk}\right) = \mu_{lk}^{*}$

$$\text{Let } I_{lk}^{'} = \begin{cases} 1 \text{ if } k\in s_l' \\ 0, \text{ otherwise} \end{cases}$$

$$I_{lk}^{*} = \begin{cases} 1 \text{ if } k\in s_l \\ 0, \text{ otherwise} \end{cases}$$

Therefore

$$E_{p_1}E_{p_2}\left(\sum_{L}\sum_{k\in S_l}\frac{\mu_{lk}^{*}\alpha_{lk}}{\pi_{lk}}\right) = E_{p_1}E_{p_2}\left[\sum_{l=1}^{L}\sum_{k\in S_l'}\frac{\mu_{lk}^{*}\beta_{lk}}{\pi_{lk}}I_{lk2}\right] = E_{p_1}\left[\sum_{l=1}^{L}\sum_{k\in S_l'}\frac{\mu_{lk}^{*}\beta_{lk}}{\pi_{lk}}\pi_{lk2}\right]$$

$$= E_{p_1}\left[\sum_{k=1}^{L}\sum_{k\in S_l'}\frac{\mu_{lk}^{*}\beta_{lk}}{\pi_{lk1}}\right] = E_{p_1}\left[\sum_{l=1}^{L}\sum_{k=1}^{N_l}\frac{\mu_{lk}^{*}\beta_{lk}}{\pi_{lk1}}I_{lk1}\right] = \sum_{l=1}^{L}\sum_{k=1}^{N_l}\mu_{lk}^{*}\beta_{lk} \qquad (16)$$

To get the variance consider

$$V\left(\hat{Y}_{GMHT}\right) = V_{p_1 p_2} m\left(\hat{Y}_{GMHT}\right)$$

$$= V_{p_1}E_{p_2}E_m\left(\hat{Y}_{GMHT}\right) + E_{p_1}V_{p_2}E_m\left(\hat{Y}_{GMHT}\right) + E_{p_1}E_{p_2}V_m\left(\hat{Y}_{GMHT}\right)$$

Now,

$$V_{p_1}E_{p_2}E_m\left(\hat{Y}_{GMHT}\right) = V_{p_1}E_{p_2}\left[\sum_{L}\sum_{k\in S_l}\frac{\mu_{lk}^{*}\beta_{lk}}{\pi_{lk}}\right] = V_{p_1}\left[\sum_{L}\sum_{k=1}^{N_l}\frac{\mu_{lk}^{*}\beta_{lk}}{\pi_{lk1}}I_{lk1}\right]$$

$$= \sum_{L}\left[\sum_{k=1}^{N_l}\frac{\mu_{lk}^{*2}\beta_{lk}^{2}}{\pi_{lk1}}\left(1-\pi_{lk1}\right) + \sum_{k\neq k'}\frac{\mu_{lk}^{*}\mu_{lk'}^{*}\beta_{lk}\beta_{lk'}}{\pi_{lk1}\pi_{lk'1}}\left(\pi_{lkk'1}-\pi_{lk1}\pi_{lk'1}\right)\right] + \sum_{l\neq l'}\sum_{k=1}^{N_l}\sum_{k'=1}^{N_{l'}}\frac{\mu_{lk}^{*}\mu_{l'k'}^{*}\beta_{lk}\beta_{l'k'}}{\pi_{lk1}\pi_{l'k'1}}\left(\pi_{ll'kk'1}-\pi_{lk1}\pi_{l'k'1}\right) \qquad (17)$$

Also

$$E_{p_1} V_{p_2} E_m = E_{p_1} V_{p_2} \left[\sum_{l=1}^{L} \sum_{k \in S_l} \frac{\mu_{lk}^* \beta_{lk}}{\pi_{lk}} \right] = E_{p_1} V_{p_2} \left[\sum_{l=1}^{L} \sum_{k \in S_{l'}} \frac{\mu_{lk}^* \beta_{lk}}{\pi_{lk}} I_{lk}'' \right]$$

$$= E_{p_1} \left\{ \sum_{l=1}^{L} \left[\sum_{k \in S_l} \frac{\mu_{lk}^{*2} \beta_{lk}^2}{\pi_{lk}^2} \pi_{lk2} \left(1 - \pi_{lk2} \right) + \sum_{k \neq k' \in S_{l'}} \frac{\mu_{lk}^*}{\pi_{lk}} \frac{\mu_{lk'}^*}{\pi_{lk'}} \beta_{lk} \beta_{lk'} \left(\pi_{lkk'2} - \pi_{lk2} \pi_{lk'2} \right) \right] \right.$$

$$\left. + \sum_{l \neq l'} \sum_{k \in S_l} \sum_{k' \in S_{l'}} \frac{\mu_{lk}^* \mu_{l'k'}^* \beta_{lk} \beta_{l'k'}}{\pi_{lk} \pi_{l'k'}} \left(\pi_{ll'kk'2} - \pi_{lk2} \pi_{l'k'2} \right) \right\}$$

$$= E_{p_1} \left\{ \sum_{l=1}^{L} \left[\sum_{k=1}^{N_l} \frac{\mu_{lk}^{*2} \beta_{lk}^2}{\pi_{lk}^2} \pi_{lk2} \left(1 - \pi_{lk2} \right) I_{lk} + \sum_{k \neq k' \in S_{l'}} \frac{\mu_{lk}^* \mu_{lk'}^* \beta_{lk} \beta_{lk'}}{\pi_{lk} \pi_{lk'}} \left(\pi_{lkk'2} - \pi_{lk2} \pi_{lk'2} \right) I_{lkk'} \right] \right.$$

$$\left. + \sum_{l \neq l'} \sum_{k \in S_l} \sum_{k \in S_{l'}} \frac{\mu_{lk}^* \mu_{l'k'}^* \beta_{lk} \beta_{l'k'}}{\pi_{lk} \pi_{l'k'}} \left(\pi_{ll'kk'2} - \pi_{lk2} \pi_{l'k'2} \right) I_{ll'kk'}' \right\}$$

$$= \sum_{l=1}^{L} \left[\sum_{k=1}^{N_l} \frac{\mu_{lk}^{*2} \beta_{lk}^2}{\pi_{lk}^2} \left(1 - \pi_{lk2} \right) I_{lk} + \sum_{k \neq k'} \frac{\mu_{lk}^* \mu_{lk'}^* \beta_{lk} \beta_{lk'}}{\pi_{lk} \pi_{lk'}} \pi_{lkk'2} \left(\pi_{lkk'2} - \pi_{lk2} \pi_{lk'2} \right) \right]$$

$$+ \sum_{l \neq l'} \sum_{k=1}^{N_l} \sum_{k'=1}^{N_{l'}} \frac{\mu_{lk}^* \mu_{l'k'}^* \beta_{lk} \beta_{l'k'}}{\pi_{lk} \pi_{l'k'}} \pi_{ll'kk'2} \left(\pi_{ll'kk'2} - \pi_{lk2} \pi_{ll'kk'} \right) \tag{18}$$

Finally,

$$E_{p_1} E_{p_2} V_m \left(\hat{Y}_{GMHT} \right) = E_1 E_2 \left\{ \sum_{l=1}^{L} \left[\sum_{k \in S_l} \frac{\beta_{lk}^2 \sigma_{lk}^2}{\pi_{lk}} + \sum_{k \neq k' \in S_l} \frac{\beta_{lk} \beta_{lk'} \sigma_{lkk'}}{\pi_{lk} \pi_{lk'}} \right] + \sum_{l \neq l'} \sum_{k \in S_l} \sum_{k' \in S_{l'}} \frac{\beta_{lk} \beta_{l'k'}}{\pi_{lk} \pi_{l'k'}} \sigma_{ll'kk'} \right\}$$

$$= E_{p_1} \left\{ \sum_{l=1}^{L} \left[\sum_{k \in S_{l'}} \frac{\beta_{lk}^2 \sigma_{lk}^2}{\pi_{lk}} \pi_{lk2} + \sum_{k \neq k' \in S_{l'}} \frac{\beta_{lk} \beta_{lk'} \sigma_{lkk'}}{\pi_{lk} \pi_{l'k}} \pi_{lkk'2} \right] + \sum_{l \neq l'} \sum_{k \in S_l} \sum_{k \in S_{l'}} \frac{\beta_{lk} \beta_{l'k'} \sigma_{ll'kk'}}{\pi_{lk} \pi_{l'k'}} \pi_{ll'kk'} \right\}$$

$$= \sum_{l=1}^{L} \left[\sum_{k=1}^{N_l} \frac{\alpha_{kl}^2 \sigma_{kl}^2}{\pi_{kl}} + \sum_{k \neq k'} \frac{\beta_{lk} \beta_{lk'} \sigma_{lkk'}}{\pi_{lk} \pi_{lk'}} \pi_{lkk'} \right] + \sum_{l \neq l'} \sum_{k=1}^{N_l} \sum_{k'=1}^{N_{l'}} \frac{\beta_{lk} \beta_{l'k'}}{\pi_{lk} \pi_{l'k'}} \sigma_{ll'kk'} \pi_{ll'kk'} \tag{19}$$

Where

$$\beta_{lk} = \frac{\pi_{lk, k \in d}}{\sum_{d=1}^{D} \sum_{l=1}^{L} \pi_{lk, k \in d}}$$

And $\pi_{lk, k \in d}$ implies that it is considered only when k belongs to domain d for all the frames.

If we consider simple random sampling then;

$$V \left(\hat{Y}_{GMHT} \right) = V_{p_1} E_{p_2} E_m \left(\hat{Y}_{GMHT} \right) + E_{p_1} V_{p_2} E_m \left(\hat{Y}_{GMHT} \right) + E_{p_1} E_{p_2} V_m \left(\hat{Y}_{GMHT} \right)$$

Where

$$V_{p_1} E_{p_2} E_m \left(\hat{Y}_{GMHT} \right) = \sum_{l=1}^{L} \left[\sum_{k=1}^{N_l} \frac{\mu_{lk}^{*2} \beta_{lk}^2}{\frac{n_l}{N_l}} \left(1 - \frac{n_l}{N_l} \right) + \sum_{k \neq k'} \sum \frac{\mu_{lk}^* \mu_{lk'}^* \beta_{lk} \beta_{lk'}}{\frac{n_l}{N_l} \cdot \frac{n_l}{N_l}} \left(\frac{n_l (n_l - 1)}{N_l (N_l - 1)} - \frac{n_l}{N_l} \cdot \frac{n_l}{N_l} \right) \right]$$

$$+ \sum_{l \neq l'} \sum \sum_{k=1}^{N_l} \sum_{k'=1}^{N_{l'}} \frac{\mu_{lk}^* \mu_{l'k'}^* \beta_{lk} \beta_{l'k'}}{\frac{n_l}{N_l} \cdot \frac{n_l}{N_l}} \left(\frac{n_l (n_l - 1)}{N_l (N_l - 1)} - \frac{n_l}{N_l} \cdot \frac{n_l}{N_l} \right)$$

$$= \sum_{l=1}^{L} \left[\left(\frac{N_l^2 (N_l - n_l)}{N^2 n_l} \right) \sum_{k=1}^{N_l} \mu_{lk}^{*2} + \frac{n_l (n_l - N_l)}{N^2 (N_l - 1)} \sum_{k \neq k'} \sum \mu_{lk}^* \mu_{lk'}^* \right] + \sum_{l \neq l'} \sum \frac{N_{l'} n_l (n_l - N_l)}{N^2 N_l (N_l - 1)} \sum_{k=1}^{N_l} \sum_{k'=1}^{N_{l'}} \mu_{lk}^* \mu_{l'k'}^* \quad (20)$$

also

$$E_{p_1} V_{p_2} E_m \left(\hat{Y}_{GMHT} \right) = \sum_{l=1}^{L} \left[\sum_{k=1}^{N_l} \frac{\mu_{lk}^{*2} \beta_{lk}^2}{\frac{n_l}{N_l}} \left(1 - \frac{n_l}{N_l} \right) + \sum_{k \neq k'} \sum \frac{\mu_{lk}^* \mu_{lk'}^* \beta_{lk} \beta_{lk'}}{\frac{n_l}{N_l} \cdot \frac{n_l}{N_l}} \frac{n_l (n_l - 1)}{N_l (N_l - 1)} \left(\frac{n_l (n_l - 1)}{N_l (N_l - 1)} - \frac{n_l}{N_l} \cdot \frac{n_l}{N_l} \right) \right]$$

$$+ \sum_{l \neq l'} \sum \sum_{k=1}^{N_l} \sum_{k'=1}^{N_{l'}} \frac{\mu_{lk}^* \mu_{l'k'}^* \beta_{lk} \beta_{l'k'}}{\frac{n_l}{N_l} \cdot \frac{n_l}{N_l}} \left(\frac{n_l (n_l - 1)}{N_l (N_l - 1)} \right) \left(\frac{n_l (n_l - 1)}{N_l (N_l - 1)} - \frac{n_l}{N_l} \cdot \frac{n_l}{N_l} \right)$$

$$= \sum_{l=1}^{L} \left[\left(\frac{N_l^2 (N_l - n_l)}{N^2 n_l} \right) \sum_{k=1}^{N_l} \mu_{lk}^{*2} + \frac{n_l^2 (n_l - N_l)(n_l - 1)}{N^2 N_l (N_l - 1)^2} \sum_{k \neq k'} \sum \mu_{lk}^* \mu_{lk'}^* \right] + \sum_{l \neq l'} \sum \frac{N_{l'} n_l^2 (n_l - N_l)(n_l - 1)}{N^2 N_l^2 (N_l - 1)^2} \sum_{k=1}^{N_l} \sum_{k'=1}^{N_{l'}} \mu_{lk}^* \mu_{l'k'}^* \quad (21)$$

and

$$E_{p_1} E_{p_2} V_m \left(\hat{Y}_{GMHT} \right) = \sum_{l=1}^{L} \left[\sum_{k=1}^{N_l} \frac{\beta_{lk}^2 \sigma_{lk}^2}{\frac{n_l}{N_l}} + \sum_{k \neq k'} \sum \frac{\beta_{lk} \beta_{lk'} \sigma_{lkk'}}{\frac{n_l}{N_l} \cdot \frac{n_l}{N_l}} \left(\frac{n_l (n_l - 1)}{N_l (N_l - 1)} \right) \right] + \sum_{l \neq l'} \sum \sum_{k=1}^{N_l} \sum_{k'=1}^{N_{l'}} \frac{\beta_{lk} \beta_{l'k'} \sigma_{lkk'}}{\frac{n_l}{N_l} \cdot \frac{n_l}{N_l}} \left(\frac{n_l (n_l - 1)}{N_l (N_l - 1)} \right)$$

$$= \sum_{l=1}^{L} \left[\frac{N_l^3}{N^2 n_l} \sum_{l=1}^{N_l} \sigma_{lk}^2 + \frac{N_l^3 (n_l - 1)}{N^2 n_l (N_l - 1)} \sum_{k \neq k'} \sum \sigma_{lkk'} \right] + \sum_{l \neq l'} \sum \frac{N_l^2 N_{l'} (n_l - 1)}{N^2 n_l (N_l - 1)} \sum_{k=1}^{N_l} \sum_{k'=1}^{N_{l'}} \sigma_{ll'kk'} \quad (22)$$

Suppose dual frames are considered under simple random sampling with measurement errors, then using the Kalton Anderson (1986) multiplicity adjustment factor we get

$$\hat{y}_{kA} = \sum_{k \in S_A} \frac{y_k}{\pi_k} \left[I_{k \in a(A)} + \frac{\pi_k}{\pi_k + \pi_k} I_{k \in a(AB)} \right] + \sum_{k \in S_B} \frac{y_k}{\pi_k} \left[I_{k \in b(B)} + \frac{\pi_k}{\pi_k + \pi_k} I_{k \in b(AB)} \right]$$

$$= \sum_{k \in S_A} \frac{y_k}{\pi_k} \left[I_{k \in a(A)} + \frac{1}{2} I_{k \in a(AB)} \right] + \sum_{k \in S_B} \frac{y_k}{\pi_k} \left[I_{k \in b(B)} + \frac{1}{2} I_{k \in b(AB)} \right]$$

$$= \frac{N}{n} \sum_{k \in S_A} y_k \left[I_{k \in a(A)} + \frac{1}{2} I_{k \in a(AB)} \right] + \frac{N}{n} \sum_{k \in S_B} y_k \left[I_{k \in b(B)} + \frac{1}{2} I_{k \in b(AB)} \right] \quad (23)$$

$$E(\hat{y}_k) = E_{pm}(\hat{y}_k)$$

$$= E_p \frac{N}{n} \sum_{k \in s_A} \mu_k^* \left[I_{k \in a(A)} + \frac{1}{2} I_{k \in a(AB)} \right] + E_p \frac{N}{n} \sum_{k \in s_b} \mu_k^* \left[I_{k \in b(B)} + \frac{1}{2} I_{k \in b(AB)} \right]$$

where $\mu_k^* = E_m(y_k)$

$$E(y_{kA}) = \frac{N}{n} \sum_{k \in s_A} \mu_k^* \left[E_p(I_{k \in a(A)}) + \frac{1}{2} E_p(I_{k \in a(AB)}) \right] + \frac{N}{n} \sum_{k \in s_b} \mu_k^* \left[E_p(I_{k \in b(B)}) + \frac{1}{2} E_p(I_{k \in b(AB)}) \right]$$

$$= \frac{N}{n} \sum_{k \in s_A} \mu_k^* \left[\pi_{k \in a(A)} + \frac{1}{2} \pi_{k \in a(AB)} \right] + \frac{N}{n} \sum_{k \in s_b} \mu_k^* \left[\pi_{k \in b(B)} + \frac{1}{2} \pi_{k \in b(AB)} \right]$$

$$V(\hat{y}_k) = V_{pm} \, m(\hat{y}_k) = V_p E_m(\hat{y}_k) + E_p V_m(\hat{y}_k)$$

$$V_p E_m(\hat{y}_{kA}) = V_p \left\{ \frac{N}{n} \sum_{k \in s_A} \mu_k^* \left[I_{k \in a(A)} + \frac{1}{2} I_{k \in a(AB)} \right] + \frac{N}{n} \sum_{k \in s_b} \mu_k^* \left[I_{k \in b(B)} + \frac{1}{2} I_{k \in b(AB)} \right] \right\}$$

$$= \frac{N^2}{n^2} \left[\begin{array}{l} \sum_{k \in s_A} \mu_k^{*2} (\pi_{k \in a(A)} - \pi_{k \in a(A)}^2) + \frac{1}{4} (\pi_{k \in a(AB)} - \pi_{k \in a(AB)}^2) + \\ \frac{1}{2} \sum_{k \in a(A)} \sum_{\neq k \in a(AB)} (\pi_{k \in a(A) k \in a(AB)} - \pi_{\in a(A)} \pi_{\in a(AB)}) \end{array} \right]$$

$$+ \frac{N^2}{n^2} \left[\sum_{k \in s_B} \mu_k^{*2} (\pi_{k \in b(B)} - \pi_{k \in b(B)}^2) + \frac{1}{4} (\pi_{k \in b(AB)} - \pi_{k \in b(AB)}^2) + \frac{1}{2} \sum_{k \in b(B)} \sum_{\neq k \in b(AB)} (\pi_{k \in b(B) k \in b(AB)} - \pi_{k \in b(B)} \pi_{k \in b(AB)}) \right] \quad (24)$$

$$E_p V_m(\hat{y}_k) = V_m \left[\frac{N}{n} \sum_{k \in s_A} y_k (I_{k \in a(A)} + \frac{1}{2} I_{k \in a(AB)}) + \frac{N}{n} \sum_{k \in s_B} y_k (I_{k \in b(B)} + \frac{1}{2} I_{k \in b(AB)}) \right]$$

$$= \frac{N^2}{n^2} \left[\sum_{k \in a(A)} V(e_k) + \frac{1}{2} \sum_{k \in a(A)} \sum_{\neq k \in a(AB)} \mathrm{Cov}(e_{k \in a(A)}, e_{k \in a(AB)}) \right] + \frac{N^2}{n^2} \left[\sum_{k \in b(B)} V(e_k) + \frac{1}{2} \sum_{k \in b(B)} \sum_{\neq k \in b(AB)} \mathrm{Cov}(e_{k \in b(B)}, e_{k \in b(AB)}) \right]$$

$$= \frac{N^2}{n^2} \left[\sum_{k \in a(A)} \sigma_{lk}^2 + \frac{1}{4} \sum_{k \in a(A)} \sum_{\neq k \in a(AB)} \sigma_{k \in a(A) k \in a(AB)} \right] + \frac{N^2}{n^2} \left[\sum_{k \in b(B)} \sigma_{lk}^2 + \frac{1}{4} \sum_{k \in b(B)} \sum_{\neq k \in b(AB)} \sigma_{k \in b(B) k \in b(AB)} \right] \quad (25)$$

6. Empirical Results

A normal population of size N=1000 was randomly simulated with mean=60 and variance=10. The population values were then assigned to the profiles randomly. The results are as shown below,

Estimate $\hat{\mu} = 60.0550$ and Estimate $Var(\hat{\mu}) = 9.8058$

A sample of size n=500 is then randomly selected from the N=1000 simulated values to obtain $\mu_i, i=1,2,...,n$. A population with measurement and response errors is then generated as $y_i, i=1,2,...,n$, where $y_i = \mu_i + e_i, i=1,2,...,n$ and $e_i \sim N(k\mu_i, p\mu_i)$, where k and p are positive constants. The population with measurement and response errors is randomly assigned to the domains for different k and p, The sample sizes

for the domains are, $n = 216$, $n_2 = 218$ and $n_3 = 227$. The results are obtained as shown in table 1.

Table 1. *Mean and variance when there are measurement and response errors.*

P	k	$\hat{\mu}$	Estimate ($\hat{\theta}$)	Var ($\hat{\mu}$)	Estimate Var ($\hat{\theta}$)
0	0	60.2571		14.4347	
	0.05		62.9512		13.2997
0.01	0.1		66.1512		13.4362
	0.2		72.6264		16.4282
	0.5		89.9158		23.8238
	0.05		63.2000		17.1787
0.05	0.1		66.3623		16.9830
	0.2		72.3944		21.4795
	0.5		90.5796		26.7637

In table 1, the mean and variance increase as the value of k and p increases. This indicates that the more the measurement and response errors the more the mean and variance are overestimated. This implies that for an increasing error (k) on the mean level, the measurement and response errors consequently increase.

6.1. Two Phase Multiple Frames

A sample of size n=500 is then randomly selected from the N=1000 simulated values as the first phase units to obtain μ_i, $i=1,2,...,n$. A population with measurement and response errors is then generated as y_i, $i=1,2,...,n$, where $y_i = \mu_i + e_i$, $i=1,2,...,n$ and $e_i \sim N(k\mu_i, p\mu_i)$, where k and p are positive constants. A sub sample of size n=200 is then selected from the sample of 500 to obtain second phase units. The second phase units with measurement and response errors is randomly assigned to the domains for different k and p. The sample sizes for the domains are, $n_1 = 150$, $n_2 = 105$ and $n_3 = 144$. The results are obtained as shown in table 2.

Table 2. *Mean and variance when there are measurement and response errors.*

P	k	$\hat{\mu}$	Estimate ($\hat{\theta}$)	Var ($\hat{\mu}$)	Estimate Var ($\hat{\theta}$)
0	0	59.6984		10.7458	
	0.05		63.3164		12.4250
	0.1		66.3837		14.1045
0.01	0.2		72.3424		16.1986
	0.5		90.4231		25.4393
	0.05		63.3347		11.2118
	0.1		66.3455		12.3914
0.05	0.2		72.2840		15.1299
	0.5		90.4749		25.5433

In table 2, the mean and variance are overestimated as the values of k and p increases. Although two phase sampling is applied to reduce the cost of surveying, sometimes the errors can be more, compared to a single phase sampling. This means that the estimates can be compromised in the presence of measurement and response errors.

6.2. Multiple Frames Using Hovitz Thompson Estimator

In the empirical example, simulation was carried out with the assumption of three frames and hence resulting in seven domains A, B, C, AB, AC, BC and ABC. The effect of errors was studied by simulating the errors as percentages of the individual true values. The model used is, $y_i = \mu_i + e_i$, $i=1,2,...,n$ and, $e_i \sim N(k\mu_i, p\mu_i)$ where k and p are constants With the assumption that there are measurement and response errors in domains the mean and the variances vary for various error percentages as below, for varying values of β_{lk}.

Table 3. *Mean and variance when there are measurement and response errors.*

P	k	$\hat{\mu}$	Estimate ($\hat{\theta}$)	Var ($\hat{\mu}$)	Estimate Var ($\hat{\theta}$)
0	0	60.4403		11.8001	
	0.05		63.0278		9.8794
	0.1		66.3158		10.0613
0.01	0.2		72.7136		12.6846
	0.5		90.0096		18.5659
	0.05		63.1338		11.6359
	0.1		66.3750		11.6865
0.05	0.2		72.4286		15.2172
	0.5		90.4635		20.5422

In table 3 the mean and variance increases as the values of k and p increases. The presence of measurement and response errors lead to overestimates in mean and variances. Comparing table 3 to table 1, the Hovitz Thompson estimator gives better estimates.

6.3. Two Phase Multiple Frames Using Hovitz Thompson Estimator

We now consider $\beta_{lk} = \frac{1}{5}$ for domain A, B and C, and $\beta_{lk} = \frac{4}{5}$ for domain AB, BC, AC and ABC

Table 4. *Mean and variance when there are measurement and response errors.*

P	k	$\hat{\mu}$	Estimate ($\hat{\theta}$)	Var ($\hat{\mu}$)	Estimate Var ($\hat{\theta}$)
0	0	59.6754		11.1655	
	0.05		63.1021		10.9621
	0.1		66.2250		11.3204
0.01	0.2		72.0380		13.4313
	0.5		90.1399		20.2481
	0.05		63.1952		12.7641
	0.1		66.2123		13.0332
0.05	0.2		72.0136		16.2225
	0.5		90.3652		22.5368

In table 4 the effect of the errors is the presence of a bias as can be seen from the elevated estimates of the population mean. Also, the presence of measurement and response errors result in overestimation of the variances.

7. Conclusion

The main aim of this work was to study the effect of measurement errors in surveys that where complete sampling frames do not exist. The process of carrying out any survey is expensive in terms of financial resources, time, planning and execution of set program. It is therefore of paramount importance to identify and reduce effects of the measurement and response errors so that the data obtained is reliable for use in social advancement programs. This paper has shown that the presence of measurement and response errors in elusive surveys where multiple frames are applied in sample selection has great effects on the results by over-estimating

the mean and under estimating of variance of the surveyed data. Also considered is the effect on the generalized-multiplicity adjusted Horvitz-Thompson estimator where there is over-estimating of the mean and under estimating of variance of the surveyed data.

References

[1] Rueda, M. D. M., Arcos, A., Molina, D., & Ranalli, M. G. (2018). Estimation techniques for ordinal data in multiple frame surveys with complex sampling designs. International Statistical Review, 86 (1), 51-67.

[2] Singh, A. C., Mecatti, F. (2014) Estimation in Multiple Frame Surveys: A simplified and unified review using Multiplicity Approach. *Journal of Official Statistics*, 155 (4), 51-69.

[3] del Mar Rueda, M., Ranalli, M. G., Arcos, A., & Molina, D. (2020). Population empirical likelihood estimation in dual frame surveys. Statistical Papers, 1-18.

[4] Iacan, D., Iachan, R., Dennis, M. L. (1993) A multiple frame approach to sampling the homeless and transient population. Journal of Official Statistics, 9, 747-764.

[5] Brukelles. (2000) Push and pull factors of international migration, country report: Italy. Technical report 5, Eurostat.

[6] Singh, A. C., Mecatti, F. (2009) A generalized multiplicity-adjusted Horvitz Thompson class of Multiple frame estimators. *Book of Abstract. ITACOSM09, June 10-12/2009, Siena, Italy*, Invited paper, 75-77.

[7] Singh, A. C., Mecatti, F. (2011) Generalized multiplicity-adjusted Horvitz Thompson type Estimation as a unified Approach to Multiple Frame Surveys. *Journal of Official Statistics*, 27 (4), 633.

[8] Kalton, G., Anderson, D. W. (1986) Sampling rare populations. *Journal of Royal Statistical Society*, A, 149 (1), 65-82.

[9] Lohr, S. L. (2009) Multiple frame survey in sample survey: Design, methods and applications. *Handbook of statistics, Pjeffermann, D., Rao, C. R. Eds*, 29, A, 71-88.

[10] Lohr, S. L. (2011) Alternative survey sample desgns : Sampling with multiple overlapping frames. *Survey Methodology*, 37 (2), 197-213.

[11] Lohr, S., Rao, J. N. K. (2006) Estimation in multiple frame surveys. *Journal of the American Statistical Association*, 101 (475), 1019-1030.

[12] Rao, J. N. K., Wu, C. (2010) Pseudo empirical likelihood inference for multiple frame surveys. Journal of the American Statistical Association, 105 (492), 1494-1503.