Two-stage informative cluster sampling - estimation and prediction

Abdulhakeem Eideh[1]
Alquds University, Palestine

Gad Nathan
Hebrew University of Jerusalem, Israel

Ray Chambers

University of Southampton, UK

Summary. The aims of this article are twofold: first estimate the parameters of the superpopulation model for two-stage cluster sampling from a finite population when the sampling design for the both stages is informative. Second predict the finite population total, cluster-specific effects for clusters in the sample and clusters not in the sample, and predict the cluster totals for clusters in the sample and clusters not in the sample. To achieve this we derive the sample and complement-sample distributions and the moments of the first stage and second stage measurements. Also we derive the conditional sample and conditional sample-complement distributions and the moments of the cluster-specific effects given the cluster measurements.

Keywords: Analysis of variance; Empirical best linear unbiased predictor; Informative sampling; Pseudo maximum likelihood; Sample distribution.

1. Introduction

Two-stage cluster sampling is frequently used in health and social sciences. Classical theory underlying the use of this sampling mechanism involves simple random sampling for each of the two stages or unequal probabilities of selection at one or more of the two stages- see Cochran (1977), Sarndal, Swensson, and Wretman (1997), and Lohr (1999) for discussion and examples. In such cases the relationship between the response variable and the covariates in the sample is the same as modeled for the population. When the selection probabilities are related to the values of the response variable even after conditioning on concomitant variables included in the population model, we obtain what is known as informative sampling, which results in selection bias, consequently the relationship between the response variable and the covariates in the sample differs from the population model, so that standard estimates of the population model parameters severely biased, leading possibly to false inference-for more discussion, see Pfeffermann, Krieger and Rinott (1998).
In a recent master thesis Amin (2000) consider the estimation of the variance components model parameters when the sampling design for the first stage is informative with exponential sampling while for the second stage is noninformative. In (2001) Pfeffermann, Moura and Silva estimate the parameters of the two-level model under informative sampling design for both stages using the Markov Chain Monte Carlo algorithm. The authors noted that the sample models used for this experiment are correct. In practice the relationship between the sample selection probabilities and the model dependent variables need to be identified from the sample.

[1] *Address for correspondence*: Department of Mathematics, Faculty of Science and Technology, Alquds University, Abu-Dies Campus, Palestine , P.O. Box 20002, East Jerusalem,

Neither of the above studies considers maximum likelihood estimation method, prediction of finite population total, cluster-specific effects for sample and non-sample clusters, and prediction of cluster totals for sample and non-sample clusters when the sampling design for any of the stages is informative.
The aims of the present study are twofold: estimation and prediction when the sampling design for the both stages is informative. Special cases are those in which sampling at one or neither of the stages is informative. In order to do that we extract the sample distributions of the first and second stages using the population model and first order inclusion probabilities, extract the sample marginal and sample posterior distributions, and then find sample and complement-sample moments of these distributions. We can thereby use the resulting sample, complement-sample and their moments for estimation of the population model parameters and prediction of finite population total, cluster-specific effects for sample and non-sample clusters, and prediction of cluster totals for sample and non-sample clusters when the sampling design for both stages is informative.
In Section 2 we define the population model for two-stage cluster sampling. Section3 defines the sampling designs. In section 4 we extract the sample pdf's for the first and second stages. In section 5 we derive the marginal and posterior sample distributions and their moments. In Section 6 we study the variance components model for two-stage informative cluster sampling. Section 7 is devoted to the estimation of population parameter and prediction of sample and non-sample cluster-specific effects. In Section 8 we predict the finite population total under-two stage informative cluster sampling. In Section 9 we predict the sample and non-sample cluster totals, and Section 10 for simulation study (later).

2. Population model and sampling design for two-stage cluster sampling

Consider a finite population U of N primary sampling units (psu's) or clusters labelled $1,...,N, U=\{1,...,N\}$ where N is a known number. Let $M_i, i=1,...,N$ be the number of secondary sampling units (ssu's) in the *ith* psu. Let $y_{ij}, i=1,...,N,\ j=1,...,M_i$ be the value of the response variable y for the ssu j belonging to the psu i. Assume that cluster-specific auxiliary data $\mathbf{z}_i=\left(z_{i1},...,z_{iq}\right)'$ and element-specific auxiliary data $\mathbf{x}_{ij}=\left(x_{ij1},...,x_{ijp}\right)'$ are available for all clusters and population elements, respectively.
Consider the following two-stage superpopulation model that includes random-intercept effects:

$$\begin{aligned} &First\text{ - }Stage: \mu_i=\mathbf{z}_i'\gamma+\eta_i \\ &Second-Stage: y_{ij}\mid\mu_i=\mu_i+\mathbf{x}_{ij}'\beta+e_{ij}, \end{aligned} \tag{1}$$

where $i=1,...,N;\ j=1,...,M_i;\ e_{ij}\underset{p}{\sim}N\left(0,\sigma_e^2\right);\ \eta_i\underset{p}{\sim}N\left(0,\sigma_\mu^2\right),\ \beta=\left(\beta_1,...,\beta_p\right)'$ and

$\gamma=\left(\gamma_1,...,\gamma_q\right)'$ are vectors of unknown fixed regression parameters. We assume that all the e_{ij}'s and η_i's are independent. Here the cluster-specific effects μ_i are modeled as linear functions of known auxiliary variables $\mathbf{z}_i$. The first-stage cluster-specific effects η_i accounts for the variation of the random cluster-specific effects not explained by the repressor variables $\mathbf{z}_i$.

For this model we have:

$$E_p\left(y_{ij}\right)=\mathbf{z}_i'\gamma+\mathbf{x}_{ij}'\beta,\;\; Var_p\left(y_{ij}\right)=\sigma_\mu^2+\sigma_e^2,\;\; Cov_p\left(y_{ij},y_{ik}\right)=\sigma_\mu^2,\; j\neq k$$
$$Cov_p\left(y_{ij},y_{rk}\right)=0,\, i\neq r \tag{2}$$

The purpose of this study is to estimate the population parameters $\gamma, \beta, \sigma_e^2$, and σ_μ^2, predicting the cluster-specific effects for clusters in the sample and clusters not in the sample, the prediction of the finite population total, and the prediction of cluster totals for sample and non-sample clusters.

Particular cases of model (1) are model, which depend on the data available, and are widely used in small area estimation, for example:

1. Random effects model or variance components model: for this model we assume that no auxiliary variables are available for both stages:

$$y_{ij}=\mu_i+e_{ij},e_{ij}\underset{p}{\sim}NI\left(0,\sigma_e^2\right)$$
$$\mu_i=\mu+\eta_i,\eta_i\underset{p}{\sim}NI\left(0,\sigma_\mu^2\right) \tag{3}$$

In matrix notation this model can be expressed as:

$$\mathbf{y}_i=\mu\mathbf{1}_{M_i}+\mathbf{u}_i,\, E_p\left(\mathbf{u}_i\right)=\mathbf{0}, E_p\left(\mathbf{u}_i\mathbf{u}_i'\right)=Cov_p\left(\mathbf{u}_i\right)=\mathbf{V}_i=\sigma_\mu^2\mathbf{J}_{M_i}+\sigma_e^2\mathbf{I}_{M_i} \tag{4}$$

where $\mathbf{1}_{M_i}=\left(1,...,1\right)'$ is a vector of length M_i and consequently $\mathbf{y}_i\underset{p}{\sim}N\left(\mu\mathbf{1}_{M_i},\mathbf{V}_i\right)$.

This random effects model is proposed by Scott and Smith (1969) as a superpopulation model for two-stage cluster sampling from a finite population.

2. Nested error unit level regression model: in this case element-specific auxiliary data are available for all population elements:

$$y_{ij}=\mathbf{x}_{ij}'\beta+\mu_i+e_{ij},e_{ij}\underset{p}{\sim}NI\left(0,\sigma_e^2\right)$$
$$\mu_i=\mu+\eta_i,\eta_i\underset{p}{\sim}NI\left(0,\sigma_\mu^2\right) \tag{5}$$

This model employed originally by Battese et al. (1988).

3. Area level random effects model: in this model cluster-specific auxiliary data are available for all clusters:

$$y_{ij}=\mu_i+e_{ij},e_{ij}\underset{p}{\sim}NI\left(0,\sigma_e^2\right)$$
$$\mu_i=\mathbf{z}_i'\gamma+\eta_i,\eta_i\underset{p}{\sim}NI\left(0,\sigma_\mu^2\right), e_{ij}\text{ and }\eta_{\text{i}}\text{ are independent} \tag{6}$$

4. The model used by Fay and Herriot (1979) is defined by:

$$\tilde{\mu}_i=\mu_i+e_i,e_i\underset{p}{\sim}NI\left(0,\sigma_{Di}^2\right)$$
$$\mu_i=\mathbf{z}_i'\gamma+\eta_i,\eta_i\underset{p}{\sim}NI\left(0,\sigma_\mu^2\right), e_i\text{ and }\eta_{\text{i}}\text{ are independent} \tag{7}$$

where $\tilde{\mu}_i$ is the direct sample estimate of μ_i.

We assume two-stage cluster sampling design with informative sampling for the first and second stages. Special cases are those in which sampling at only one or neither of the stages is informative.

Let $d_i, d_{ij}, i=1,...,N,\; j=1,...,M_i$ be design variables (considered as random), used for the sample selection but not included in the working model under consideration.At

the first stage a sample s of size n psu's (clusters) is selected with inclusion probabilities:
$\pi_i = \Pr(i \in s \mid \mu_i, \mathbf{z}_i, d_i) = h_1(\mu_i, \mathbf{z}_i, d_i)$ for some function h_1 and all psu's $i \in U$.
At the second stage a sample, s_i, of size m_i ssu's is selected from the $i'th$ selected psu with conditional inclusion probabilities:
$\pi_{j|i} = \Pr(j \in s_i | i \in s, y_{ij}, \mathbf{x}_{ij}, d_{ij}) = h_2(y_{ij}, \mathbf{x}_{ij}, d_{ij})$ for some function h_2, for all ssu's $j \in U_i$ and all psu's $i \in U$.
In fact these first order selection probabilities are a function of observed auxiliary variables which are related to μ_i and y_{ij}. For simplicity we consider the first order inclusion probabilities to depend on the unobserved μ_i and y_{ij} instead of the actual variables for selection, which are related to them.
In the following we use only the conditional expectations of the inclusion probabilities $E_p(\pi_i \mid \mu_i, \mathbf{z}_i)$ and $E_p(\pi_{j|i} \mid \mathbf{x}_{ij}, y_{ij})$.
Comment 1: If $\pi_{j|i}$=1, then we are in single-stage cluster sampling.

3. Sample model and sample moments for two-stage informative cluster sampling

In this section we will derive the sample distributions and their moments for the cluster-specific effects μ_i (first stage), for $y_{ij} \mid \mu_i$ (second stage), and derive their moments.

3.1 Sample distribution and sample moments for the cluster-specific effects

Following Pfeffermann, Krieger and Rinott (1998), the first stage sample distribution of the cluster-specific effects μ_i is:

$$f_s(\mu_i \mid \mathbf{z}_i) = \frac{E_p(\pi_i \mid \mu_i, \mathbf{z}_i)}{E_p(\pi_i \mid \mathbf{z}_i)} f_p(\mu_i \mid \mathbf{z}_i) \qquad (8)$$

where $E_p(\pi_i \mid \mathbf{z}_i) = \int E_p(\pi_i \mid \mu_i, \mathbf{z}_i) f_p(\mu_i \mid \mathbf{z}_i) d\mu_i$.
Note that the sample distribution is different from the population distribution unless $E_p(\pi_i \mid \mu_i, \mathbf{z}_i) = E_p(\pi_i \mid \mathbf{z}_i)$ for all $i \in U$. In such cases the sampling design is ignorable. Also the sample distribution of μ_i is fully determined by the conditional expectation $E_p(\pi_i \mid \mu_i, \mathbf{z}_i)$ and the superpopulation model distribution $f_p(\mu_i \mid \mathbf{z}_i)$.
Suppose the sample inclusion probabilities have an approximate expectation:

A. Exponential model: $$E_p(\pi_i \mid \mu_i, \mathbf{z}_i) = g_e(\mathbf{z}_i)\exp(b_0 + b_1\mu_i) \qquad (9a)$$

B. Linear model: $$E_p(\pi_i \mid \mu_i, \mathbf{z}_i) = a_0 + a_1\mu_i + g_l(\mathbf{z}_i) \qquad (9b)$$

where g_l and g_e are some functions of $\mathbf{z}_i$ and $\{a_0, a_1, b_0, b_1\}$ are unknown parameters to be estimated from the sample data (see the section on estimation) .
Under the exponential model, we can show that, the sample pdf of μ_i is given by:

$$f_s(\mu_i \mid \mathbf{z}_i) = (2\pi\sigma_\mu^2)^{-0.5} \exp\left(-\frac{1}{2\sigma_\mu^2}\left(\mu_i - (\mathbf{z}_i'\gamma + b_1\sigma_\mu^2)\right)^2\right) \qquad (10)$$

Hence the sample and population models belong to the normal distribution, but the mean of the sample model shifts by the constant $b_1\sigma^2_\mu$, so that

$$E_s(\mu_i \mid \mathbf{z}'_i) = \mathbf{z}'_i\gamma + b_1\sigma^2_\mu \text{ and } Var_s(\mu_i \mid \mathbf{z}'_i) = \sigma^2_\mu \tag{11}$$

Note that the sample pdf of μ_i is independent of b_0 and. If $b_1 = 0$, then the sample and population distributions are the same, in such cases the sampling mechanism for the second stage is ignorable.

Under the linear model, the sample pdf of μ_i is given by:

$$f_s(\mu_i \mid \mathbf{z}_i) == A_0 f_p(\mu_i \mid \mathbf{z}_i) + A_1 f_p^w(\mu_i \mid \mathbf{z}_i) \tag{12}$$

where $A_0 = \dfrac{a_0 + g_l(\mathbf{z}_i)}{a_0 + a_1\mathbf{z}'_i\gamma + g_l(\mathbf{z}_\mathrm{i})}$, $A_1 = 1 - A_0$ and $f_p^w(\mu_\mathrm{i} \mid \mathbf{z}_\mathrm{i}) = \dfrac{\mu_i f_p(\mu_\mathrm{i} \mid \mathbf{z}_\mathrm{i})}{E_p(\mu_i \mid \mathbf{z}_i)}$.

That is, $f_s(\mu_i \mid \mathbf{z}_\mathrm{i})$ is a mixture of normal and weighted normal distribution of μ_i given $\mathbf{z}_i$ in the population.

Note that the sample and population pdf's are different unless $a_1 = 0$, in which case the sampling mechanism for the first stage is noninformative. Also if $a_1 \neq 0$, $g_l(\mathbf{z}_i) = 0$, and $a_0 = 0$, then $f_s(\mu_i \mid \mathbf{z}_\mathrm{i}) = f_p^w(\mu_i \mid \mathbf{z}_\mathrm{i})$, the weighted distribution of μ_i.

Now, we can show thatthe mean and the variance under the sample pdf of μ_i in (12) are:

$$E_s(\mu_i \mid \mathbf{z}'_i) = \mathbf{z}'_i\gamma + \frac{a_1\sigma^2_\mu}{(\mathrm{a}_0 + g_l(\mathbf{z}_i)) + a_1\mathbf{z}'_i\gamma} \tag{13a}$$

$$Var_s(\mu_i \mid \mathbf{z}'_i) = \sigma^2_\mu\left(1 - \frac{(a_1\sigma_\mu)^2}{(a_0 + g_l(\mathbf{z}_i) + a_1\mathbf{z}'_i\gamma)^2}\right) \tag{13b}$$

Here the mean and the variance are changing. Thus according to (10) and (12) the change in the population parameters depends on the model of the conditional expectation: $E_p(\pi_i \mid \mu_\mathrm{i}, \mathbf{z}_i)$. Also note that, according to (13bb), $Var_s(\mu_i \mid \mathbf{z}'_i) \leq Var_p(\mu_i \mid \mathbf{z}'_i)$ with equality if and if $a_1 = 0$, that is when the sampling design is noninformative.

3.2 Sample distribution and sample moments of second stage measurements

Similarly to Section 3.1, the conditional sample pdf of y_{ij} given μ_i and $\mathbf{x}_{ij}$ is given by:

$$f_s(y_{ij} \mid \mu_i, \mathbf{x}_{ij}) = \frac{E_p(\pi_{j|i} \mid \mu_i, \mathbf{x}_{ij}, y_{ij})}{E_p(\pi_{j|i} \mid \mu_i, \mathbf{x}_{ij})} f_p(y_{ij} \mid \mu_i, \mathbf{x}_{ij}) \tag{14}$$

Consider the following model assumption for the conditional expectation of the sample selection probabilities.

A. Exponential model: $E_p(\pi_{j|i} \mid \mathbf{x}_{ij}, y_{ij}, \mu_i) = k_e(\mathbf{x}_{ij})\exp(d_0 + d_1 y_{ij})$ (15a)

B. Linear model: $E_p(\pi_{j|i} \mid \mathbf{x}_{ij}, y_{ij}, \mu_i) = c_0 + c_1 y_{ij} + k_l(\mathbf{x}_{ij})$ (15b) where k_l and k_e are some functions of $\mathbf{x}_{ij}$ and μ_i and $\{d_0, d_1, c_0, c_1\}$ are unknown parameters to be estimated from the sample data (see the section on estimation).

Then under the exponential model, the conditional sample pdf of y_{ij} given μ_{i} is given by:

$$f_s\left(y_{ij} \mid \mu_i, \mathbf{x}_{ij}\right) = \left(2\pi\sigma_e^2\right)^{-0.5} \exp\left(-\frac{1}{2\sigma_e^2}\left(y_{ij} - \left(\mu_i + \mathbf{x}'_{ij}\beta + d_1\sigma_e^2\right)\right)^2\right) \quad (16)$$

Also, $E_s\left(y_{ij} \mid \mu_i\right) = \mu_i + \mathbf{x}'_{ij}\beta + d_1\sigma_e^2$ and $Var_s\left(\mu_i \mid \mathbf{z}'_i\right) = \sigma_e^2$.
Similarly under the linear model we have:

$$f_s\left(y_{ij} \mid \mu_{\mathrm{i}}, \mathbf{x}_{ij}\right) = \frac{\left(c_0 + k_l\left(\mathbf{x}_{ij}\right) + c_1 y_{ij}\right)}{c_0 + k_l\left(\mathbf{x}_{ij}\right) + c_1\mu_i + c_1\mathbf{x}'_{ij}\beta} f_p\left(y_{ij} \mid \mu_i, \mathbf{x}_{ij}\right) \quad (16)$$

The mean and the variance under this sample model are:

$$E_s\left(y_{ij} \mid \mu_i, \mathbf{x}_{ij}\right) = \mu_i + \mathbf{x}'_{ij}\beta + \frac{c_1\sigma_e^2}{\left(c_0 + k_l\left(\mathbf{x}_{ij}\right)\right) + c_1\left(\mu_i + \mathbf{x}'_{ij}\beta\right)} \quad (17a)$$

$$Var_s\left(y_{ij} \mid \mu_i, \mathbf{x}_{ij}\right) = \sigma_e^2\left(1 - \frac{\left(c_1\sigma_e^2\right)^2}{\left(\left(c_0 + k_l\left(\mathbf{x}_{ij}\right)\right) + c_1\left(\mu_i + \mathbf{x}'_{ij}\beta\right)\right)^2}\right) \quad (17b)$$

4. Sample marginal distributions of cluster measurements

In this section we are interested in deriving the sample marginal distribution of $\mathbf{y}_i = \left(y_{i1}, \ldots, y_{im_i}\right)$ and the conditional sample distribution of $\mu_i \mid \mathbf{y}_i$, and their moments and complement-sample moments, when the sampling design for the first stage and second stage is informative. Special cases are those in which sampling at one or neither of the stages is informative. Under this sampling mechanism, we showed that the sample and population distributions for the first and second stages are different-see Section 3. The sample model and sample moments of y_{ij} for the two-stage depend on the sample model of the cluster-specific effects (first stage) and the sample model of $y_{ij} \mid \mu_{\mathrm{i}}$ (second stage). So that we have different combinations to consider, depending on the modelling of the conditional expectation assumed for the first order inclusion probabilities of selecting clusters in the first stage and selecting units in the second stage from the selected clusters in the first stage. As an illustration let us consider the following two cases:

Case1: Model (9a) $E_p\left(\pi_i \mid \mu_{\mathrm{i}}, \mathbf{z}_i\right) = g_e\left(\mathbf{z}_i\right)\exp\left(b_0 + b_1\mu_i\right)$ and

Model (15a) $E_p\left(\pi_{j|i} \mid y_{ij}, \mathbf{x}_{\mathrm{ij}}, \mu_i\right) = k_e\left(\mathbf{x}_{ij}\right)\exp\left(d_0 + d_1 y_{ij}\right)$

Using the results of Sections 3.1and 3.2 and method of conditioning, we can show that:

The mean of y_{ij} under the sample model is:

$$E_s\left(y_{ij}\right) = \mathbf{z}'_i\gamma + \mathbf{x}'_{ij}\beta + d_1\sigma_e^2 + b_1\sigma_\mu^2 \quad (18a)$$

The variance of y_{ij} under the sample model is:

$$Var_s\left(y_{ij}\right) = \sigma_e^2 + \sigma_\mu^2 \quad (18b)$$

The covariance between y_{ij} and y_{ik} under the sample model is:

$$Cov_s(y_{ij}, y_{ik}) = \sigma_\mu^2 \tag{18c}$$

Note that in this case the mean changes, while the variance and covariance do not change.

For this case let us find the marginal sample pdf of $(y_{i1},..., y_{im_i})$ .Using (10), (16) and since $(y_{i1},..., y_{im_i}) | \mu_i, i = 1,..., n$ are asymptotically independent, then we can show that:

$$f_s(\mathbf{y}_i) = (2\pi)^{-0.5m_i} (\sigma_e^2)^{-0.5(m_i-1)} (m_i\sigma_\mu^2 + \sigma_e^2)^{-0.5}$$

$$\exp\left(-\frac{1}{2\sigma_e^2}\left(\sum_{j=1}^{m_i}\left(y_{ij} - \left(\mathbf{x}'_{ij}\beta + \mathbf{z}'_i\gamma + b_1\sigma_\mu^2 + d_1\sigma_e^2\right)\right)^2\right)\right)* \tag{19}$$

$$\exp\left(\frac{\sigma_\mu^2}{(2\sigma_e^2)(m_i\sigma_\mu^2 + \sigma_e^2)}\left(\sum_{j=1}^{m_i}\left(y_{ij} - \left(\mathbf{x}'_{ij}\beta + \mathbf{z}'_i\gamma + b_1\sigma_\mu^2 + d_1\sigma_e^2\right)\right)\right)^2\right)$$

Notice that the sample and population marginal distributions belong to the same family, but the mean under the sample model is shifts by the constant $b_1\sigma_\mu^2 + d_1\sigma_e^2$, this constant represent the effect of informative sampling for both stages.

This sample pdf's can be used for estimating the population parameters $\gamma, \beta, \sigma_\mu^2$ and σ_e^2 .

Case 2: Model (9b) $E_p(\pi_i | \mu_\mathrm{i}, \mathbf{z}_i) = a_0 + a_1\mu_i + g_l(\mathbf{z}_i)$ and

Model (15a) $E_p(\pi_{j|i} | y_{ij}, \mathbf{x}_\mathrm{ij}, \mu_i) = k_e(\mathbf{x}_{ij})\exp(d_0 + d_1 y_{ij})$

Similar to Case 1, we can show that:

$$E_s(y_{ij}) = \mathbf{x}'_{ij}\beta + \mathbf{z}'_i\gamma + d_1\sigma_e^2 + \frac{a_1\sigma_\mu^2}{a_0 + g_l(\mathbf{z}_i) + a_1\mathbf{z}'_i\gamma} \tag{20a}$$

$$Var_s(y_{ij}) = \sigma_e^2 + \sigma_\mu^2\left(1 - \frac{(\mathrm{a}_1\sigma_\mu)^2}{(a_0 + g_l(\mathbf{z}_i) + a_1\mathbf{z}'_i\gamma)^2}\right) \tag{20b}$$

and

$$Cov_s(y_{ij}, y_{ik}) = \sigma_\mu^2\left(1 - \frac{(a_1\sigma_\mu)^2}{(a_0 + g_l(z_i) + a_1\mathbf{z}'_i\gamma)^2}\right) \tag{20c}$$

Also we can show that:

$$f_s(\mathbf{y}_i) = h_s(\mathbf{y}_i)\left(A_0 + \frac{A_1}{\mathbf{z}'_i\gamma}\left(\frac{\sigma_\mu^2\sum_{j=1}^{m_i}(y_{ij} - \mathbf{x}'_{ij}\beta - d_1\sigma_e^2) + \sigma_e^2(\mathbf{z}'_i\gamma)}{m_i\sigma_\mu^2 + \sigma_e^2}\right)\right) \tag{21}$$

where

$A_0 = \dfrac{a_0 + g_l(\mathbf{z}_i)}{a_0 + g_l(\mathbf{z}_i) + a_1(\mathbf{z}'_i\gamma)}$, $A_1 = 1 - A_0$ and $h_s(\mathbf{y}_i)$ is the marginal distribution of $\mathbf{y}_i$ when the sampling design for the first stage is noninformative and for the second stage is informative, which is the marginal distribution given in (19) with $b_1 = 0$.

5. Estimation of population parameters

We are interested in estimating the vector of unknown population parameters, $\theta_p = \left(\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma_\mu^2, \sigma_e^2\right)$, that characterize the population models given in (1). We base the inference on the sample models derived in Section 4. We consider two methods of estimation, the two-step and pseudo maximum likelihood estimation methods.

5.1 Two-setp method:

Case 1: Model (9a) $E_p\left(\pi_i \mid \mu_i, \mathbf{z}_i\right) = \exp\left(b_1\mu_i + \mathbf{b}'\mathbf{z}_i\right)$ and

Model (15a) $E_p\left(\pi_{j|i} \mid y_{ij}, \mathbf{x}_{ij}, \mu_i\right) = \exp\left(d_1 y_{ij} + \mathbf{d}'\mathbf{x}_{ij}\right)$

where

$\mathbf{b} = \left(b_0, b_2, ..., b_q\right)'$, $\mathbf{z}_i = \left(1, z_{i1}, ..., z_{iq}\right)'$, $\mathbf{d} = \left(d_0, d_2, ..., d_p\right)'$ and $\mathbf{x}_{ij} = \left(1, x_{ij1}, ..., x_{ijp}\right)'$.

Under this approximation, some of the parameters of the sample model are not identifiable, so that we estimate the parameters of the population model based on the sample models in two steps.

Step-one: the coefficients b_1 and $\mathbf{b}$ and d_1 and $\mathbf{d}$ are estimated via the relationship between the moments under the sample and population models- see Pfeffermann and Sverchkov (1999). Thus we have:

$$E_s\left(w_i \mid \mu_i\right) = \exp\left(-\left(b_1\mu_i + \mathbf{b}'\mathbf{z}_i\right)\right) \tag{22a}$$

and

$$E_s\left(w_{j|i} \mid \mu_i\right) = \exp\left(-\left(d_1 y_{ij} + \mathbf{d}'\mathbf{x}_{ij}\right)\right) \tag{22b}$$

where $w_i = \pi_i^{-1}$ and $w_{j|i} = \pi_{j|i}^{-1}$.

The problem in estimating b_1 and $\mathbf{b}$ is that, the μ_i are unobserved, one possible solution is to estimate μ_i by $\bar{y}_i = m_i^{-1}\sum_{j=1}^{m_i} y_{ij}$ and use measurement error model. Another possibility for estimating b_1 and $\mathbf{b}$ is to replace μ_i in (22a) by $\bar{y}_i$. Thus the OLS estimators of $\mathbf{B} = \left(-b_1, -\mathbf{b}'\right)'$ is given by:

$$\tilde{\mathbf{B}} = \left(-\tilde{b}_1, -\tilde{\mathbf{b}}'\right)' = \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{W}, \quad \mathbf{Z}_i = \left(\bar{y}_i, \mathbf{z}_i\right)' \tag{23}$$

where $\mathbf{Z}_i = \left(\bar{y}_i, \mathbf{z}_i\right)'$.

Similarly, using (22b), the OLS estimators of $\mathbf{D} = \left(-d_1, \mathbf{d}'\right)'$ is given by:

$$\tilde{\mathbf{D}} = \left(-\tilde{d}_1, -\tilde{\mathbf{d}}\right)' = \sum_{i=1}^{n}\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\sum_{i=1}^{n}\left(\mathbf{X}_i'\mathbf{W}_i\right) \tag{24}$$

where $\mathbf{X}_i = \left(\mathbf{y}_i', \mathbf{x}_i'\right)$ and $\mathbf{W}_i = \left(\log w_{1|i}, ..., \log w_{m_i|i}\right)$, $i = 1, ..., n$.

Step-two: having estimating $\left(b_1, d_1\right)$, we substitute these estimates in (19) and then the next step is to estimate the parameters, $\theta = \left(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_e^2, \sigma_\mu^2\right)$, of the superpopulation model given in (4.1) via the ML estimation method.

According to (19), the contribution to the log-likelihood function for the ith sampled cluster can be written as:

$$L_{rs_i}\left(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_e^2,\sigma_\mu^2\right)=-0.5(m_i-1)\log\left(\sigma_e^2\right)-0.5\log\left(m_i\sigma_\mu^2+\sigma_e^2\right)$$

$$-\frac{1}{2\sigma_e^2}\sum_{j=1}^{m_i}\left(y_{ij}-\left(\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_i\boldsymbol{\gamma}+\tilde{b}_1\sigma_\mu^2+\tilde{d}_1\sigma_e^2\right)\right)^2 \qquad (25a)$$

$$+\frac{\sigma_\mu^2}{2\left(\sigma_e^2\right)\left(m_i\sigma_\mu^2+\sigma_e^2\right)}\left(\sum_{j=1}^{m_i}\left(y_{ij}-\left(\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_i\boldsymbol{\gamma}+\tilde{b}_1\sigma_\mu^2+\tilde{d}_1\sigma_e^2\right)\right)\right)^2$$

Thus the full log-likelihood function to be maximized is given by:

$$L_{rs}\left(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_e^2,\sigma_\mu^2\right)=\sum_{i=1}^{n}L_{rs_i}\left(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_e^2,\sigma_\mu^2\right) \qquad (25b)$$

This function can be maximized using S-PLUS statistical software.

Case 2: Model (9b) $E_p\left(\pi_i \mid \mu_i,\mathbf{z}_i\right)=\left(a_1\mu_i+\mathbf{a}'\mathbf{z}_i\right)$ and

Model (15a) $E_p\left(\pi_{j|i} \mid y_{ij},\mathbf{x}_{ij},\mu_i\right)=\left(d_1 y_{ij}+\mathbf{d}'\mathbf{x}_{ij}\right)$, where

$\mathbf{a}=\left(a_0,a_2,\ldots,a_q\right)'$, $\mathbf{z}_i=\left(1,z_{i1},\ldots,z_{iq}\right)'$, $\mathbf{d}=\left(d_0,d_2,\ldots,d_p\right)'$ and $\mathbf{x}_{ij}=\left(1,x_{ij1},\ldots,x_{ijp}\right)'$.

We estimate the population model parameters in the same way as Case 1.

Step-one: Estimation of $\left(a_1,\mathbf{a}\right)$ and $\left(d_1,\mathbf{d}\right)$.

The estimates of $\left(d_1,\mathbf{d}\right)$ is given by (24), while the estimates of $\left(a_1,\mathbf{a}\right)$ are similar to (23) but $\mathbf{W}_i=\left(\frac{1}{w_{1|i}},\ldots,\frac{1}{w_{m_i|i}}\right), i=1,\ldots,n$.

Step-two: having estimating the informativeness parameters under the two stages we substitute these estimates in (21) and then the next step is to estimate the parameters, $\theta=\left(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_e^2,\sigma_\mu^2\right)$, of the superpopulation model given in (1) via the ML estimation method.

According to (21), the contribution to the log-likelihood function for the *ith* sampled cluster can be written as:

$$L_{rs_i}\left(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_e^2,\sigma_\mu^2\right)=-0.5(m_i-1)\log\left(\sigma_e^2\right)-0.5\log\left(m_i\sigma_\mu^2+\sigma_e^2\right)$$

$$-\frac{1}{2\sigma_e^2}\sum_{j=1}^{m_i}\left(y_{ij}-\left(\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_i\boldsymbol{\gamma}+\tilde{d}_1\sigma_e^2\right)\right)^2$$

$$+\frac{\sigma_\mu^2}{2\left(\sigma_e^2\right)\left(m_i\sigma_\mu^2+\sigma_e^2\right)}\left(\sum_{j=1}^{m_i}\left(y_{ij}-\left(\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_i\boldsymbol{\gamma}+\hat{d}_1\sigma_e^2\right)\right)\right)^2 \qquad (26a)$$

$$+\log(A_0+\frac{A_1}{\mathbf{z}'_i\boldsymbol{\gamma}}\left(\frac{\sigma_\mu^2\sum_{j=1}^{m_i}\left(y_{ij}-\mathbf{x}'_{ij}\boldsymbol{\beta}-\tilde{d}_1\sigma_e^2\right)+\sigma_e^2\left(\mathbf{z}'_i\boldsymbol{\gamma}\right)}{m_i\sigma_\mu^2+\sigma_e^2}\right)$$

Thus the full log-likelihood function to be maximized is given by:

$$L_{rs}\left(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_e^2,\sigma_\mu^2\right)=\sum_{i=1}^{n}L_{rs_i}\left(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_e^2,\sigma_\mu^2\right) \qquad (26b)$$

This function can be maximized using S-PLUS software.

Pseudo maximum likelihood estimation

Now we extend the idea of PML to two-stage cluster sampling. According to Section 2 , the first stage inclusion probabilities are denoted by $\pi_i, i=1,...,N$ and the second stage inclusion probabilities are denoted by $\pi_{j|i}, j=1,...,M_i$. So that the joint inclusion probabilities are given by $\pi_{ij}=\pi_i \pi_{j|i}, i=1,...,N;\ j=1,...,M_i$. Therefore the joint sampling weights are $w_{ij}=w_i w_{j|i}, w_i=\pi_i^{-1}, w_{j|i}=\pi_{j|i}^{-1}, i=1,...,N;\ j=1,...,M_i$.

Under the conditions of model (1), the contribution of the ith cluster to the census log-likelihood function is given by:

$$\begin{aligned} L_{C_i}\left(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_e^2,\sigma_\mu^2\right) &= -0.5(M_i-1)\log\left(\sigma_e^2\right)-0.5\log\left(M_i\sigma_\mu^2+\sigma_e^2\right) \\ &-\frac{1}{2\sigma_e^2}\sum_{j=1}^{M_i}\left(y_{ij}-\left(\mathbf{x}_{ij}'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\gamma}\right)\right)^2 \\ &+\frac{\sigma_\mu^2}{2\left(\sigma_e^2\right)\left(M_i\sigma_\mu^2+\sigma_e^2\right)}\left(\sum_{j=1}^{M_i}\left(y_{ij}-\left(\mathbf{x}_{ij}'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\gamma}\right)\right)\right)^2 \end{aligned} \tag{27a}$$

Thus the full census log-likelihood function to be maximized is given by:

$$L_C\left(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_e^2,\sigma_\mu^2\right)=\sum_{i=1}^{N}L_{C_i}\left(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_e^2,\sigma_\mu^2\right) \tag{27b}$$

Now the probability weighted estimator of $L_{C_i}\left(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_e^2,\sigma_\mu^2\right)$ is given by:

$$\begin{aligned} \hat{L}_{C_i}\left(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_e^2,\sigma_\mu^2\right) &= -0.5(m_i-1)\log\left(\sigma_e^2\right)-0.5\log\left(m_i\sigma_\mu^2+\sigma_e^2\right) \\ &-\frac{1}{2\sigma_e^2}\sum_{j=1}^{m_i}w_{j|i}\left(y_{ij}-\left(\mathbf{x}_{ij}'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\gamma}\right)\right)^2 \\ &+\frac{\sigma_\mu^2}{2\left(\sigma_e^2\right)\left(m_i\sigma_\mu^2+\sigma_e^2\right)}\left(\sum_{j=1}^{m_i}w_{j|i}\left(y_{ij}-\left(\mathbf{x}_{ij}'\boldsymbol{\beta}+\mathbf{z}_i'\boldsymbol{\gamma}\right)\right)\right)^2 \end{aligned} \tag{27c}$$

Thus the PML estimator is defined as the solution of following estimating equation:

$$\hat{L}_C\left(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_e^2,\sigma_\mu^2\right)=\sum_{i=1}^{n}w_i\hat{L}_{C_i}\left(\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_e^2,\sigma_\mu^2\right) \tag{27d}$$

This function can be maximized using S-PLUS software.

6. Prediction of Cluster-Specific Effects

In order to predict the cluster-specific effects for cluster in the sample and cluster not in the sample, we first derive the conditional sample distribution of $\mu_i \mid y_{i1},...,y_{im_i}$, and their moments and complement-sample moments, when the sampling design for the first stage and second stage is informative. Special cases are those in which sampling at one or neither of the stages is informative. Under this sampling mechanism, we showed that the sample and population distributions for the first and second stages are different-see Sections 3.1 and 3.2. The conditional sample model and conditional sample moments of $\mu_i \mid y_{i1},...,y_{im_i}$ depend on the sample model of the cluster-specific effects (first stage) and the sample model of $y_{ij} \mid \mu_i$ (second stage). So that we have different combinations to consider, depending on the modelling of the conditional expectation assumed for the first order inclusion probabilities of selecting clusters in the first stage and selecting units in the second stage from the selected clusters in the first stage. As an illustration let us consider the following two cases:

Case 1: Model (9a) $E_p(\pi_i \mid \mu_i, \mathbf{z}_i) = \exp(b_1\mu_i + \mathbf{b}'\mathbf{z}_i)$ and

Model (15a) $E_p(\pi_{j|i} \mid y_{ij}, \mathbf{x}_{ij}, \mu_i) = \exp(d_1 y_{ij} + \mathbf{d}'\mathbf{x}_{ij})$

In this case we can show that, the conditional sample pdf of $\mu_i \mid \mathbf{y}_i$ is given by:

$$\mu_i \mid \mathbf{y}_i \underset{s}{\sim} N\left(\frac{\sigma_\mu^2 \sum_{j=1}^{m_i}\left(y_{ij} - \mathbf{x}'_{ij}\beta - d_1\sigma_e^2\right) + \sigma_e^2\left(\mathbf{z}'_i\gamma + b_1\sigma_\mu^2\right)}{m_i\sigma_\mu^2 + \sigma_e^2}, \frac{\sigma_e^2\sigma_\mu^2}{m_i\sigma_\mu^2 + \sigma_e^2}\right) \tag{28}$$

Thus after some calculations we can show that:

$$\begin{aligned} E_s(\mu_i \mid \mathbf{y}_i) &= \frac{\sigma_\mu^2 \sum_{j=1}^{m_i}\left(y_{ij} - \mathbf{x}'_{ij}\beta - d_1\sigma_e^2\right) + \sigma_e^2\left(\mathbf{z}'_i\gamma + b_1\sigma_\mu^2\right)}{m_i\sigma_\mu^2 + \sigma_e^2} \\ &= E_p(\mu_i|\mathbf{y}_i) + (1 - \phi_i)b_1\sigma_\mu^2 - \phi_i d_1\sigma_e^2 \end{aligned} \tag{29}$$

where

$E_p(\mu_i|\mathbf{y}_i) = \phi_i\left(\bar{y}_i - m_i^{-1}\sum_{j=1}^{m_i}\mathbf{x}'_{ij}\beta\right) + (1 - \phi_i)(\mathbf{z}'_i\gamma)$ and $\phi_i = \dfrac{m_i\sigma_\mu^2}{m_i\sigma_\mu^2 + \sigma_e^2}$.

This conditional expectation can be written as:

$$E_s(\mu_i \mid \mathbf{y}_i) = \frac{Var_s(\mu_i)}{Var_s(\bar{y}_i)}\left(\bar{y}_i - E_s(\bar{y}_i)\right) + E_s(\mu_i) \tag{30}$$

where $E_s(\bar{y}_i) = \dfrac{1}{m_i}\sum_{j=1}^{m_i}\left(\mathbf{x}'_{ij}\beta + \mathbf{z}'_i\gamma + b_1\sigma_\mu^2 + d_1\sigma_e^2\right)$, $E_s(\mu_i) = \mathbf{z}'_i\gamma + b_1\sigma_\mu^2$ and

$Var_s(\bar{y}_i) = \dfrac{m_i\sigma_\mu^2 + \sigma_e^2}{m_i}$.

Also (29) can be interpreted as a weighted average of $\bar{y}_i - m_i^{-1}\sum_{j=1}^{m_i}\mathbf{x}'_{ij}\beta - d_1\sigma_e^2$ and

$\mathbf{z}'_i\gamma + b_1\sigma_\mu^2$ with weights $\dfrac{m_i\sigma_\mu^2}{m_i\sigma_\mu^2 + \sigma_e^2}$ and $\left(1 - \dfrac{m_i\sigma_\mu^2}{m_i\sigma_\mu^2 + \sigma_e^2}\right)$ respectively.

Prediction of cluster-specific effects for clusters in the sample:

Simple calculations show that for known variances $(\sigma_\mu^2, \sigma_e^2)$, and known informativeness parameters (b_1, d_1), the best linear unbiased predictor (BLUP) of the cluster-specific effects for clusters in the sample is given by:

$$\hat{\mu}_{i,s}\big|\mathbf{y}_s = \phi_i\left(\bar{y}_i - m_i^{-1}\sum_{j=1}^{m_i}\mathbf{x}'_{ij}\hat{\beta}\right) + (1-\phi_i)(\mathbf{z}'_i\hat{\gamma}) + (1-\phi_i)b_1\sigma_\mu^2 - \phi_i d_1\sigma_e^2 ,$$
$$= \hat{\mu}_{i,u}\big|\mathbf{y}_s + (1-\phi_i)(\mathbf{z}'_i\hat{\gamma}) + (1-\phi_i)b_1\sigma_\mu^2 - \phi_i d_1\sigma_e^2 \quad (31a)$$

where

$\phi_i = \dfrac{m_i\sigma_\mu^2}{m_i\sigma_\mu^2 + \sigma_e^2}$ and $\hat{\mu}_{i,u}\big|\mathbf{y}_s = \phi_i\left(\bar{y}_i - m_i^{-1}\sum_{j=1}^{m_i}\mathbf{x}'_{ij}\hat{\beta}\right) + (1-\phi_i)(\mathbf{z}'_i\hat{\gamma})$ is the BLUP under noninformative probability sampling for both stages.

In practice, the variances $(\sigma_\mu^2, \sigma_e^2)$ and the informativeness parameters (b_1, d_1) are unknown. A suggested procedure is to replace the variances $(\sigma_\mu^2, \sigma_e^2)$ in the BLUP by standard variance components estimates like ML estimators, PML estimators or ANOVA estimators and to replace the informativeness parameters (b_1, d_1) by their estimates-see Section 5. The resulting predictors are known as the Empirical BLUP (EBLUP). Thus the EBLUP of cluster-specific effects for clusters in the sample is given by:

$$\hat{\mu}_{i,s}\big|\mathbf{y}_s = \hat{\phi}_i\left(\bar{y}_i - m_i^{-1}\sum_{j=1}^{m_i}\mathbf{x}'_{ij}\hat{\beta}\right) + (1-\hat{\phi}_i)(\mathbf{z}'_i\hat{\gamma}) + (1-\hat{\phi}_i)\tilde{b}_1\sigma_\mu^2 - \hat{\phi}_i\tilde{d}_1\hat{\sigma}_e^2 \quad (31b)$$

where $\hat{\phi}_i = \dfrac{m_i\hat{\sigma}_\mu^2}{m_i\hat{\sigma}_\mu^2 + \hat{\sigma}_e^2}$.

Prediction of cluster-specific effects for clusters not in the sample:

Let us now find the complement-sample conditional mean of $\mu_j \mid \mathbf{y}_1,...,\mathbf{y}_n$ where $j \notin s$.

Sverchkov and Pfeffermann (2001) proved the following formula:

$$E_c\left(y_j\big|x_j\right) = \frac{E_p\left((1-\pi_j)y_j\big|x_j\right)}{E_p\left((1-\pi_j)\big|x_j\right)} \quad (32)$$

Using this formula, and if $E_p(\pi_j \mid \mu_{\mathrm{j}}, \mathbf{z}_j) = g_e(\mathbf{z}_j)\exp(b_0 + b_1\mu_j)$ then we can show that:

$$E_c\left(\mu_j\big|\mathbf{y}_1,...,\mathbf{y}_n\right) = \mathbf{z}'_j\gamma - \frac{g_e(\mathbf{z}_j)b_1\sigma_\mu^2\exp\left(b_0 + b_1(\mathbf{z}'_j\gamma) + 0.5b_1^2\sigma_\mu^2\right)}{1 - g_e(\mathbf{z}_j)\exp\left(b_0 + b_1(\mathbf{z}'_j\gamma) + 0.5b_1^2\sigma_\mu^2\right)}$$
$$= E_p\left(\mu_j\big|\mathbf{z}_j\right) - \frac{g_e(\mathbf{z}_j)b_1\sigma_\mu^2\exp\left(b_0 + b_1(\mathbf{z}'_j\gamma) + 0.5b_1^2\sigma_\mu^2\right)}{1 - g_e(\mathbf{z}_j)\exp\left(b_0 + b_1(\mathbf{z}'_j\gamma) + 0.5b_1^2\sigma_\mu^2\right)} \quad (33a)$$

Thus the EBLUP of cluster-specific effects for clusters not in the sample is given by:

$$\hat{\mu}_{j,c}\big|\mathbf{y}_s = \mathbf{z}'_j\hat{\gamma} - \frac{\tilde{g}_e(\mathbf{z}_j)\tilde{b}_1\hat{\sigma}_\mu^2\exp\left(\tilde{b}_0 + \tilde{b}_1(\mathbf{z}'_j\hat{\gamma}) + 0.5\tilde{b}_1^2\hat{\sigma}_\mu^2\right)}{1 - \tilde{g}_e(\mathbf{z}_j)\exp\left(\tilde{b}_0 + b_1(\mathbf{z}'_j\hat{\gamma}) + 0.5b_1^2\hat{\sigma}_\mu^2\right)} \quad (33b)$$

Remark 1: Notice that BLUP of cluster-specific effects for clusters in the sample, given in (31a), depends on the informative parameters b_1 and d_1 for both stages, while the BLUP of cluster-specific effects for clusters not in the sample given by (33a) depends only on the informative parameters b_0 and b_1 of the first stage. This is intuitively reasonable because for non-sample clusters we do not do any subsampling.

Case 2: Model (9b) $E_p\left(\pi_i \mid \mu_i, \mathbf{z}_i\right) = a_0 + a_1\mu_i + g_l\left(\mathbf{z}_i\right)$ and

Model (15a) $E_p\left(\pi_{j|i} \mid y_{ij}, \mathbf{x}_{ij}, \mu_i\right) = k_e\left(\mathbf{x}_{ij}\right)\exp\left(d_0 + d_1 y_{ij}\right)$

Similar to Case 1, using the results of Sections 4 and using Bayes theorem, we can show that, the conditional sample pdf of $\mu_i \mid \left(y_{i1},..., y_{im_i}\right)$ is given by:

$$f_s\left(\mu_i \mid \mathbf{y}_i\right) = \frac{\left(A_0 + \frac{A_1}{\mathbf{z}_i'\gamma}\mu_i\right)}{\left(A_0 + \frac{A_1}{\mathbf{z}_i'\gamma}E_h\left(\mu_i \mid \mathbf{y}_i\right)\right)} h_s\left(\mu_i \mid \mathbf{y}_i\right) \tag{34}$$

where $A_0 = \dfrac{a_0 + g_l\left(\mathbf{z}_i\right)}{a_0 + g_l\left(\mathbf{z}_i\right) + a_1\left(\mathbf{z}_i'\boldsymbol{\gamma}\right)}$, $A_1 = 1 - A_0$ and

$$h_s\left(\mu_i \mid \mathbf{y}_i\right) \text{is } N\left(\frac{\sigma_\mu^2 \sum_{j=1}^{m_i}\left(y_{ij} - \mathbf{x}_{ij}'\beta - d_1\sigma_e^2\right) + \sigma_e^2\left(\mathbf{z}_i'\gamma\right)}{m_i\sigma_\mu^2 + \sigma_e^2}, \frac{\sigma_e^2\sigma_\mu^2}{m_i\sigma_\mu^2 + \sigma_e^2}\right) \tag{35}$$

Notice that h_s is the conditional sample distribution of μ_i given $\mathbf{y}_i$ when the sampling design for the first stage is noninformative, while for the second stage is informative with exponential sampling. Also (34) can be written as a mixture of conditional sample distribution, $h_s\left(\mu_i \mid \mathbf{y}_i\right)$, and weighted conditional sample distributions, $h_s^w\left(\mu_i \mid \mathbf{y}_i\right)$, as follows:

$$f_s\left(\mu_i \mid \mathbf{y}_i\right) = A_0' h_s\left(\mu_i \mid\right) + A_1' h_s^w\left(\mu_i \mid \mathbf{y}_i\right) \tag{36}$$

where

$$A_0' = \frac{A_0}{\left(A_0 + \frac{A_1}{\mathbf{z}_i'\gamma}E_{h_s}\left(\mu_i \mid \mathbf{y}_i\right)\right)}, A_1' = 1 - A_0' \text{ and } h_s^w\left(\mu_i \mid \mathbf{y}_i\right) = \frac{\mu_i h_s\left(\mu_i \mid \mathbf{y}_i\right)}{E_{h_s}\left(\mu_i \mid \mathbf{y}_i\right)}.$$

These conditional sample pdf can be used for predicting μ_i.

Now we can show that the mean of the conditional sample distribution is:

$$E_s\left(\mu_i \mid \mathbf{y}_i\right) = E_{h_s}\left(\mu_i \mid \mathbf{y}_i\right) + \frac{A_1 Var_{h_s}\left(\mu_i \mid \mathbf{y}_i\right)}{\left(\mathbf{z}_i'\gamma\right)\left(A_0 + \frac{A_1}{\mathbf{z}_i'\gamma}E_{h_s}\left(\mu_i \mid \mathbf{y}_i\right)\right)} \tag{37}$$

where

$$\begin{aligned} E_{h_s}\left(\mu_i \mid \mathbf{y}_i\right) &= \frac{\sigma_\mu^2 \sum_{j=1}^{m_i}\left(y_{ij} - \mathbf{x}_{ij}'\beta - d_1\sigma_e^2\right) + \sigma_e^2\left(\mathbf{z}_i'\gamma\right)}{m_i\sigma_\mu^2 + \sigma_e^2} \\ &= E_p\left(\mu_i \mid \mathbf{y}_i\right) - d_1\sigma_e^2\phi_i \end{aligned} \tag{38a}$$

and

$$Var_{h_s}(\mu_i \mid \mathbf{y}_i) = \frac{\sigma_\mu^2 \sigma_e^2}{m_i \sigma_\mu^2 + \sigma_e^2} = Var_p(\mu_i \mid \mathbf{y}_i) \tag{38b}$$

Notice that $E_s(\mu_i \mid \mathbf{y}_i) = E_{h_s}(\mu_i \mid \mathbf{y}_i)$ if and only if $A_1 = 0$, and this happens when $a_1 = 0$, that is, when the sampling design for the first stage is noninformative. Also $E_{h_s}(\mu_i \mid \mathbf{y}_i) = E_p(\mu_i \mid \mathbf{y}_i)$ if and only if $d_1 = 0$, that is the sampling design for the second stage is noninformative.

Under this sample conditional pdf, simple calculations show the EBLUP of cluster-specific effects for clusters in the sample is given by:

$$\hat{\mu}_{i,s} \mid \mathbf{y}_s = \hat{E}_{h_s}(\mu_i \mid \mathbf{y}_i) + \frac{\tilde{A}_1 \hat{V}ar_{h_s}(\mu_i \mid \mathbf{y}_i)}{(\mathbf{z}_i'\hat{\gamma})\left(\tilde{A}_0 + \frac{\tilde{A}_1}{\mathbf{z}_i'\hat{\gamma}} \hat{E}_{h_s}(\mu_i \mid \mathbf{y}_i)\right)} \tag{39a}$$

where

$$\hat{E}_{h_s}(\mu_i \mid \mathbf{y}_i) = \frac{\hat{\sigma}_\mu^2 \sum_{j=1}^{m_i} \left(y_{ij} - \mathbf{x}_{ij}'\hat{\beta} - \tilde{d}_1 \hat{\sigma}_e^2\right) + \hat{\sigma}_e^2 (\mathbf{z}_i'\hat{\gamma})}{m_i \hat{\sigma}_\mu^2 + \hat{\sigma}_e^2}, \hat{V}ar_{h_s}(\mu_i \mid \mathbf{y}_i) = \frac{\hat{\sigma}_\mu^2 \hat{\sigma}_e^2}{m_i \hat{\sigma}_\mu^2 + \hat{\sigma}_e^2}$$

$$\tilde{A}_0 = \frac{\tilde{a}_0 + \tilde{g}_l(\mathbf{z}_i)}{\tilde{a}_0 + \tilde{g}_l(\mathbf{z}_i) + \tilde{a}_1(\mathbf{z}_i'\boldsymbol{\gamma})} \text{ and } \tilde{A}_1 = 1 - \tilde{A}_0 .$$

Prediction of cluster-specific effects for clusters not in the sample:

Let us now find the conditional complement-sample mean of $\mu_j \mid \mathbf{y}_1, ..., \mathbf{y}_n;\ \mathbf{y}_i = (y_{i1}, ..., y_{im_i})'$ where $j \notin s$. Similar to Case 1, but here: $E_p(\pi_j \mid \mu_j, \mathbf{z}_j) = (a_0 + g_l(\mathbf{z}_j) + a_1 \mu_j)$. In this case we can show that the EBLUP of cluster-specific effects for clusters not in the sample is given by:

$$\hat{\mu}_{j,c} \big| \mathbf{y}_s = \mathbf{z}_j'\hat{\gamma} - \frac{\tilde{a}_1 \hat{\sigma}_\mu^2}{1 - \left(\tilde{a}_0 + \tilde{g}_l(\mathbf{z}_j) + \tilde{a}_1(\mathbf{z}_j'\hat{\gamma})\right)} \tag{40}$$

Remark 1: Notice that EBLUP of cluster-specific effects for clusters in the sample, given in (39a), depends on the informative parameters a_0, a_1 and d_1 for both stages, while the EBLUP of cluster-specific effects for clusters not in the sample given by (40) depends only on the informative parameters a_0 and a_1 of the first stage. This is intuitively reasonable because for non-sample clusters we do not do any subsampling.

7. Prediction of finite population total under two-stage informative cluster sampling

Assume two-stage population model (1). Let

$$T = \sum_{i=1}^{N} \sum_{j=1}^{M_i} y_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} y_{ij} + \sum_{i=1}^{n} \sum_{j=m_i+1}^{M_i} y_{ij} + \sum_{i=n+1}^{N} \sum_{j=1}^{M_i} y_{ij} \tag{41}$$

define the population total that we want to predict using the sample data from two stages and possibly values of auxiliary variables that may contains some or all the design variables.
Notice that the population total is decomposed into three components: the first component represents the total for observed units, the second component represents the total for unobserved units in sample clusters, and the third component represent the total for non-sample clusters.
For the prediction process we have the following available information:
A. The information that comes from the first stage denoted by: $O_1 = \{(\mathbf{z}_i, I_i), i \in U, \pi_i, i \in s\}$.
B. Information that comes from the second stage denoted by: $O_2 = \{(y_{ij}, \pi_{j|i}), i \in s, j \in s_i, (\mathbf{x}_{ij}, I_{j|i}), i \in U, j \in U_i\}$.

Thus the available information, from both stages, for the prediction process is $O_s = O_1 \cup O_2$.
Let $\hat{T} = \hat{T}(O_s)$ define the predictor of T based on O_s. The mean square error (MSE) of $\hat{T}$ given O_s with respect to the population pdf is defined by:

$$\begin{aligned} MSE_P(\hat{T}) &= E_P\left((\hat{T}-T)^2 \mid O_s\right) = E_I\left[E_{II}\left((\hat{T}-T)^2 \mid \mu_i\right) \mid O_s\right] \\ &= E_P\left((\hat{T} - E_P(T \mid O_s) + E_P(T \mid O_s) - T)^2 \mid O_s\right) \\ &= (\hat{T} - E_P(T \mid O_s))^2 + Var_p(T \mid O_s) \end{aligned} \quad (42)$$

Using routine differentiation, we can show that (42) is minimized when $\hat{T} = E_I(E_{II}(T \mid \mu_i) \mid \mathrm{O_s})$, where $E_I(\cdot)$ and $E_{II}(\cdot)$ denote the expectations under the population distributions for the first and second stages, respectively. Now we consider the following:

$$E_I\left[(E_{II}(T \mid \mu_i)) \mid \mathrm{O_s}\right] = E_I\left[E_{II}\left(\left(\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij} + \sum_{i=1}^{n}\sum_{j=m_i+1}^{M_i} y_{ij} + \sum_{i=n+1}^{N}\sum_{j=1}^{M_i} y_{ij}\right) \mid \mu_i\right) \mid \mathrm{O_s}\right] \quad (43)$$

Thus the general predictor for the finite population total under two-stage informative cluster sampling is:

$$\hat{T} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij} + \sum_{i=1}^{n}\sum_{j=m_i+1}^{M_i} \hat{E}_I^s\left(\hat{E}_{II}^c(y_{ij} \mid \mu_i) \mid \mathrm{O_s}\right) + \sum_{i=n+1}^{N}\sum_{j=1}^{M_i} \hat{E}_I^c\left(\hat{E}_{II}(y_{ij} \mid \mu_i)\right) \quad (44)$$

So that, the prediction problem reduces to the prediction of, $E_I^s\left(E_{II}^c(y_{ij} \mid \mu_\mathrm{i}) \mid \mathrm{O_s}\right)$ and $E_I^c\left(E_{II}(y_{ij} \mid \mu_\mathrm{i})\right)$, where $E_I^s(\cdot), E_I^c(\cdot), E_{II}^c(\cdot)$, and $E_{II}^s(\cdot)$ denote the expectations under the sample, sample-complement distributions for the first and second stages, respectively.

The predictor given in (43) represents the prediction of the finite population total for two-stage sampling when the sampling mechanisms for the two stages are informative. We now consider the following particular cases:
1. First stage is noninformative and second stage is noninformative: here the predictor is given by:

$$\hat{T}_{nn} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij} + \sum_{i=1}^{n}\sum_{j=m_i+1}^{M_i} \hat{E}_I\left(\hat{E}_{II}\left(y_{ij} \mid \mu_i\right) \mid \mathrm{O_s}\right) + \sum_{i=n+1}^{N}\sum_{j=1}^{M_i} \hat{E}_I\left(\hat{E}_{II}\left(y_{ij} \mid \mu_i\right)\right) \qquad (45a)$$

Using (29) with $b_1 = d_1 = 0$, we can show that:

$$\begin{aligned}\hat{T}_{nn} &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij} + \sum_{i=1}^{n}\sum_{j=m_i+1}^{M_i}\left(\mathbf{x}'_{ij}\hat{\beta} + \hat{\phi}_i\left(\bar{y}_i - m_i^{-1}\sum_{j=1}^{m_i}\mathbf{x}'_{ij}\hat{\beta}\right) + \left(1-\hat{\phi}_i\right)\left(\mathbf{z}'_i\hat{\gamma}\right)\right) + \\ &\quad \sum_{i=n+1}^{N}\sum_{j=1}^{M_i}\left(\mathbf{x}'_{ij}\hat{\beta} + \mathbf{z}'_i\hat{\gamma}\right)\end{aligned} \qquad (45b)$$

Notice that in $\hat{T}_{nn}$, non-sample units in sample cluster i are predicted by, $\mathbf{x}_{ij}\hat{\beta} + \hat{\phi}_i\left(\bar{y}_i - m_i^{-1}\sum_{j=1}^{m_i}\mathbf{x}'_{ij}\hat{\beta}\right) + \left(1-\hat{\phi}_i\right)\left(\mathbf{z}'_i\hat{\gamma}\right)$ while all units in non-sample clusters are predicted by, $\mathbf{x}_{ij}\hat{\beta} + \mathbf{z}'_i\hat{\gamma}$.

This predictor was obtained and studied by Scott and Smith (1969) using Bayesian approach and by Royal (1976) using non-Bayesian approach. So here a new justification of the Royal predictor, and also a generalization from variance components model to linear mixed models, and in the following a generalization to two-stage informative cluster sampling.

2. First stage is informative and second stage is noninformative: in this case our predictor is:

$$\hat{T}_{in} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij} + \sum_{i=1}^{n}\sum_{j=m_i+1}^{M_i} \hat{E}_I^s\left(\hat{E}_{II}\left(y_{ij} \mid \mu_i\right) \mid O_s\right) + \sum_{i=n+1}^{N}\sum_{j=1}^{M_i} \hat{E}_I^c\left(\hat{E}_{II}\left(y_{ij} \mid \mu_i\right)\right) \qquad (46)$$

This predictor is fully determined by the modelling of the first order inclusion probabilities π_i. Consider the following conditional expectations:

Exponential model: $E_I\left(\pi_i \mid \mu_i, z_i\right) = \exp\left(b_0 + b_1\mu_i + \mathbf{z}'_i\mathbf{b}\right)$

Under this model of first order inclusion probabilities we can show that:

$$\begin{aligned}\hat{T}_{in}^{e} &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij} + \sum_{i=1}^{n}\sum_{j=m_i+1}^{M_i}\left(\mathbf{x}'_{ij}\hat{\beta} + \hat{\phi}_i\left(\bar{y}_i - m_i^{-1}\sum_{j-1}^{m_i}\mathbf{x}'_{ij}\hat{\beta}\right) + \left(1-\hat{\phi}_i\right)\left(\mathbf{z}'_i\hat{\gamma} + \tilde{b}_1\hat{\sigma}_\mu^2\right)\right) + \\ &\quad \sum_{i=n+1}^{N}\sum_{j=1}^{M_i}\left(\mathbf{x}'_{ij}\hat{\beta} + \mathbf{z}'_i\hat{\gamma} - \frac{\tilde{b}_1\hat{\sigma}_\mu^2 \exp\left(\tilde{b}_0 + \mathbf{z}'_i\tilde{\mathbf{b}} + \tilde{b}_1\mathbf{z}'_i\hat{\gamma} + 0.5\tilde{b}_1^2\hat{\sigma}_\mu^2\right)}{1-\exp\left(\tilde{b}_0 + \mathbf{z}'_i\tilde{\mathbf{b}} + \tilde{b}_1\mathbf{z}'_i\hat{\gamma} + 0.5\tilde{b}_1^2\hat{\sigma}_\mu^2\right)}\right)\end{aligned} \qquad (47)$$

Linear model: $E_I\left(\pi_i \mid \mu_i, z_i\right) = \left(a_0 + a_1\mu_i + \mathbf{z}'_i\mathbf{a}\right)$

Here we can show that:

$$\begin{aligned}\hat{T}_{in}^{l} &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij} + \sum_{i=1}^{n}\sum_{j=m_i+1}^{M_i}\left(\mathbf{x}'_{ij}\hat{\beta} + \hat{E}_I\left(\mu_i \middle| O_s\right) + \frac{\tilde{A}_1\hat{V}ar_I\left(\mu_i \middle| O_s\right)}{\mathbf{z}'_i\hat{\gamma}\left(\tilde{A}_0 + \dfrac{\tilde{A}_1\hat{E}_I\left(\mu_i \middle| O_s\right)}{\mathbf{z}'_i\hat{\gamma}}\right)}\right) + \\ &\quad \sum_{i=n+1}^{N}\sum_{j=1}^{M_i}\left(\mathbf{x}'_{ij}\hat{\beta} + \mathbf{z}'_i\hat{\gamma} - \frac{\tilde{a}_1\hat{\sigma}_\mu^2}{1-\left(\tilde{a}_0 + \tilde{a}_1\mathbf{z}'_i\hat{\gamma} + \mathbf{z}'_i\tilde{\mathbf{a}}\right)}\right)\end{aligned} \qquad (48)$$

where

$$E_I\left(\mu_i\middle|O_s\right)=\phi_i\left(\bar{y}_i-m_i^{-1}\sum_{j=1}^{m_i}\mathbf{x}'_{ij}\beta\right)+\left(1-\phi_i\right)\left(\mathbf{z}'_i\gamma\right),\ Var_I\left(\mu_i\middle|O_s\right)=\frac{\sigma_e^2\sigma_\mu^2}{m_i\sigma_\mu^2+\sigma_e^2}=\frac{\phi_i\sigma_e^2}{m_i}$$

$A_0=\dfrac{a_0+\mathbf{z}'_i\mathbf{a}}{a_0+\mathbf{z}'_i\mathbf{a}+a_1\left(\mathbf{z}'_i\gamma\right)}$ and $A_1=\dfrac{a_1\left(\mathbf{z}'_i\gamma\right)}{a_0+\mathbf{z}'_i\mathbf{a}+a_1\left(\mathbf{z}'_i\gamma\right)}$.

Other cases: first stage noninformative and second stage informative and first and second stages informative can be treated in the same way but in these cases predictors have no closed forms so Taylor approximation or Monte Carlo simulation are adopted.

8. Prediction of finite population totals for sampled and non-sampled clusters

In Section 8 we obtained predictors of the finite population total. But in some applications like small area estimation we are interested in predicting the small area totals for sampled and non-sampled small areas.

Assume two-stage population model (5.1). Let

$$T_i=\sum_{j=1}^{M_i}y_{ij} \tag{49}$$

define the population total for cluster $i\in U$ that we want to predict using the sample data from two stages and possibly values of auxiliary variables that may contains some or all the design variables.

Prediction of cluster totals for clusters in the sample:

In order to do that, let us decompose T_i into two components: the first component represents the total for observed units, and the second component represents the total for unobserved units in sample clusters that is:

$$T_i=\sum_{j=1}^{M_i}y_{ij}=\sum_{j=1}^{m_i}y_{ij}+\sum_{j=m_i+1}^{M_i}y_{ij}=\sum_{j\in s_i}y_{ij}+\sum_{j\notin s_i}y_{ij} \tag{50}$$

Let $\hat{T}_i\left(O_s\right)$ denotes the predictor of T_i based on O_s. So as in Section 8, the mean square error with respect to the population model of $\hat{T}_i$ is minimised when $\hat{T}_i=E_I\left(E_{II}\left(T_i\mid\mu_i\right)\mid O_s\right)$. Thus we have:

$$E_I\left[\left(E_{II}\left(T_i\mid\mu_i\right)\right)\mid O_s\right]=\sum_{j=1}^{m_i}y_{ij}+\sum_{j=m_i+1}^{M_i}E_I\left(E_{II}\left(y_{ij}\mid\mu_i\right)\mid O_s\right) \tag{51}$$

Thus the general predictor for the finite cluster total under two-stage informative cluster sampling is:

$$\hat{T}_i=E_I\left[\left(E_{II}\left(T_i\mid\mu_i\right)\right)\mid O_s\right]=\sum_{j=1}^{m_i}y_{ij}+\sum_{j=m_i+1}^{M_i}\hat{E}_I^s\left(\hat{E}_{II}^c\left(y_{ij}\mid\mu_i\right)\mid O_s\right) \tag{52}$$

So that, the prediction problem reduces to the prediction of, $E_I^s\left(E_{II}^c\left(y_{ij}\mid\mu_i\right)\mid O_s\right)$ under a specified sampling design.

1. First stage is noninformative and second stage is noninformative: here the predictor is given by:

$$\hat{T}_{nn}(i \in s) = \sum_{j=1}^{m_i} y_{ij} + \sum_{j=m_i+1}^{M_i} \left(\mathbf{x}'_{ij}\hat{\beta} + \hat{\phi}_i \left(\bar{y}_i - m_i^{-1} \sum_{j=1}^{m_i} \mathbf{x}'_{ij}\hat{\beta} \right) + \left(1-\hat{\phi}_i\right)\left(\mathbf{z}'_i \hat{\gamma}\right) \right) \quad (53)$$

Notice that in $\hat{T}_{nn}(i \in s)$, non-sample units in sample cluster i are predicted by, $\mathbf{x}_{ij}\hat{\beta} + \hat{\phi}_i \left(\bar{y}_i - m_i^{-1} \sum_{j=1}^{m_i} \mathbf{x}'_{ij}\hat{\beta} \right) + \left(1-\hat{\phi}_i\right)\left(\mathbf{z}'_i \hat{\gamma}\right)$.

2. First stage is informative and second stage is noninformative: in this case if the conditional expectation of the first order inclusion probabilities is exponential, then

$$\hat{T}^{e}_{in}(i \in s) = \sum_{j=1}^{m_i} y_{ij} + \sum_{j=m_i+1}^{M_i} \left(\mathbf{x}'_{ij}\hat{\beta} + \hat{\phi}_i \left(\bar{y}_i - m_i^{-1} \sum_{j-1}^{m_i} \mathbf{x}'_{ij}\hat{\beta} \right) + \left(1-\hat{\phi}_i\right)\left(\mathbf{z}'_i \hat{\gamma} + \tilde{b}_1 \hat{\sigma}^2_{\mu}\right) \right) (54)$$

Now if the conditional expectation of the first order inclusion probabilities is linear, then we have:

$$\hat{T}^{l}_{in}(i \in s) = \sum_{j=1}^{m_i} y_{ij} + \sum_{j=m_i+1}^{M_i} \left(\mathbf{x}'_{ij}\hat{\beta} + \hat{E}_I\left(\mu_i | O_s\right) + \frac{\tilde{A}_1 \hat{Var}_I\left(\mu_i | O_s\right)}{\mathbf{z}'_i\hat{\gamma}\left(\tilde{A}_0 + \frac{\tilde{A}_1 \hat{E}_I\left(\mu_i | O_s\right)}{\mathbf{z}'_i\hat{\gamma}} \right)} \right) \quad (55)$$

where

$$E_I\left(\mu_i | O_s\right) = \phi_i \left(\bar{y}_i - m_i^{-1} \sum_{j=1}^{m_i} \mathbf{x}'_{ij}\beta \right) + \left(1-\phi_i\right)\left(\mathbf{z}'_i\gamma\right),\ Var_I\left(\mu_i | O_s\right) = \frac{\sigma_e^2 \sigma_\mu^2}{m_i \sigma_\mu^2 + \sigma_e^2} = \frac{\phi_i \sigma_e^2}{m_i}$$

and

$$A_0 = \frac{a_0 + \mathbf{z}'_i\mathbf{a}}{a_0 + \mathbf{z}'_i\mathbf{a} + a_1\left(\mathbf{z}'_i\gamma\right)} \text{ and } A_1 = \frac{a_1\left(\mathbf{z}'_i\gamma\right)}{a_0 + \mathbf{z}'_i\mathbf{a} + a_1\left(\mathbf{z}'_i\gamma\right)}.$$

Similar procedures for other cases.

Prediction of cluster totals for clusters not in the sample

Here the decomposition of the finite cluster total into observed and unobserved units does not help, because for $i \notin s$ we do not observe any unit. Thus similar to the previous section, the predictor of the non-sample finite cluster total is:

$$\hat{T}_i(i \notin s) = E_I\left[\left(E_{II}\left(T_i \mid \mu_i\right)\right)\right] = \sum_{j=1}^{M_i} \hat{E}^c_I\left(\hat{E}_{II}\left(y_{ij} \mid \mu_i\right)\right) \quad (56)$$

Let us study this predictor under different sampling design for the first stage, and since we do not observe any unit in the second stage therefore no sampling design is considered of the second stage.

1. First stage noninformative: Under this sampling design our predictor is given by:

$$\hat{T}_{nn}(i \notin s) = \sum_{j=1}^{M_i} \left(\mathbf{x}'_{ij}\hat{\beta} + \mathbf{z}'_i\hat{\gamma} \right) \quad (57)$$

2. First stage informative:

Under the exponential model we have the following predictor:

$$\hat{T}^{e}_{in}(i \notin s) = \sum_{j=1}^{M_i} \left(\mathbf{z}'_i\hat{\gamma} + \mathbf{x}'_{ij}\hat{\beta} - \frac{\tilde{b}_1 \hat{\sigma}^2_\mu \exp\left(\tilde{b}_0 + \mathbf{z}'_i\tilde{\mathbf{b}} + \tilde{b}_1\left(\mathbf{z}'_i\hat{\gamma}\right) + 0.5\tilde{b}_1^2\hat{\sigma}^2_\mu\right)}{1 - \exp\left(\tilde{b}_0 + \mathbf{z}'_i\tilde{\mathbf{b}} + \tilde{b}_1\left(\mathbf{z}'_i\hat{\gamma}\right) + 0.5\tilde{b}_1^2\hat{\sigma}^2_\mu\right)} \right) \quad (58)$$

Under the linear model, the predictor is given by:

$$\hat{T}_{in}^{l}\left(i \notin s\right) = \sum_{j=1}^{M_i}\left(\mathbf{z}_i'\hat{\gamma} + \mathbf{x}'_{ij}\hat{\beta} - \frac{\tilde{a}_1\hat{\sigma}_{\mu}^{2}}{1-\left(\tilde{a}_0 + \mathbf{z}_i'\tilde{\mathbf{a}} + \tilde{a}_1\left(\mathbf{z}_i'\hat{\gamma}\right)\right)}\right) \tag{59}$$

References

Amin, M. (2001). *Estimation of the parameters of a two-stage model under informative sampling, based on the sampling distribution*, Master thesis, Department of Statistics, Hebrew University of Jerusalem.

Chambers, R. and Skinner, C. (2003).*Analysis of Survey Data*. New York: John Wiley.

Cochran, W.G. (1977). *Sampling Techniques* (3rd Ed.). New York: John Wiley.

Lohr, S.L. (1999). *Sampling: design and analysis*, London: Brooks/Cole.

Pfeffermann, D., Krieger, A.M, and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica* 8: 1087-1114.

Pfeffermann, D., Moura, F. A. S. and Nascimaneto-Silva, P.L. (2001). Multilevel modeling under informative sampling. *Proceedings of the Invited Paper Sessions Organized by the International Association of Survey Statisticians (IASS) for the 53rd session of the International Statistical Institute*, pp. 505-532.

Pfeffermann, D., Skinner, C.J., Goldstein, H., Holmes, D.J, and Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models (with discussion). *Journal of the Royal Statistical Society*, *Series B*, 60: 23-40.

Pfeffermann, D. and Sverchkov, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhya*, 61, Ser. B, Pt. 1: 166-186

Royall, R.M. (1976). The linear least squares prediction approach to two-stage sampling. *Journal of the American Statistical Association* 71: 657-664.

Sarndal, C-E., Swensson, B., and Wretman, J. (1992). *Model assisted survey sampling*, New York: Springer.

Skinner, C.J. (1994). Sample models and weights. *American Statistical Association Proceeding of the Section on Survey Research Methods*, 133-142.

Statistical Sciences, (1990). *S-Plus Reference Manual*, Seattle: Statistical Sciences.

Sverchkov, M. and Pfeffermann, D. (2001). Prediction of finite population totals under informative sampling utilizing the sample distribution. *American Statistical Association Proceedings of the Section on Survey Research Methods*, 41-46.