Design Consistent Estimators for a Mixed Linear Model on Survey Data

Rong Huang and Mike Hidiroglou

Business Survey Methods Division, Statistics Canada, Ottawa, Ontario K1Y 0A6

ABSTRACT

Some investigations associated with large government surveys typically require statistical analysis for populations that have a complex hierarchical structure. Classical analysis often fail to account for the nature of complex sampling designs and possibly result in incorrect inference for the parameters of interest. Linear mixed models can be used to analyze survey data collected from such populations in order to incorporate the complex hierarchical design structure. In this paper, we develop a method for estimating the parameters of the linear mixed model accounting for such sampling designs. We obtain the pseudo best linear unbiased estimators for the fixed and random effects by solving weighted sample estimating equations. The use of survey weights results in design consistent estimation. We also derive estimators for variance components for the nested error linear regression model. We compare the efficiency of the proposed estimators with that of existing estimators using a simulation study. This simulation study uses a two stage sampling design. Several informative or non-informative sampling schemes are considered in the simulation.

1 Introduction

Many large government surveys conducted by government agencies utilize multistage sampling schema because of the nature of the multilevel or hierarchical structure of the population from which the sample units are selected. Consequently, investigations associated with such surveys typically require statistical analysis for populations that have a complex hierarchical structure.

Classical analysis often fail to account for the nature of complex sampling designs and possibly result in biased estimators for the parameters of interest. Linear mixed models (LMM) or multilevel models (Goldstein, 1995) have been used to analyze survey data collected from populations with complex hierarchical design structures. Related work can be found in Henderson (1975), Laird and Ware (1982), Battese, et.al. (1983), Prasad and Rao (1990), and Schall (1991). Ignoring sampling designs might still be able to produce unbiased estimators for the parameter of interest in some cases, such as, when the characteristics of the sampling design can be fully explained by the LMM (non-informative design). However, when sampling designs are informative in the sense that sampling weights depend on the values of the response, even after conditioning on the covariates in the LMM, the conventional estimators of the model parameters can be biased. How to incorporate informative sampling designs into the multilevel analysis has generated much research, especially in the field of small area estimation (Prasad and Rao (1999), and You and Rao (2002)). Analytical methods for estimating model parameters incorporating informative designs are desired for general LMM on complex survey data.

In this paper, we develop a general method for estimating the parameters of the linear mixed model on survey data accounting for sampling designs. In section 2, we introduce the LMM we have considered and two special cases that have been widely employed for analyzing survey data. In section 3, we propose the pseudo empirical best linear unbiased estimators for the fixed and random effects by solving weighted sample estimating equations (WSEE). In section 4, we develop estimators for variance components for the nested error linear regression model. In section 5, we discuss the model-based and design-based properties of the estimators for variance components. We then compare the efficiency of the proposed estimators with that of existing estimators using a simulation study in section 6. We conclude our paper in section 7.

2 Model

Suppose that a population U has a two-level structure represented by levels 1 and 2. Using the Goldstein (1995) terminology, level 2 has M clusters and the i^{th} cluster has N_i level 1 elements. Let Y_{ij} denote the response of interest from the j^{th} element in the i^{th} cluster. Let $\mathbf{z}_{ij} = vec\{z_{kij}\}_{k=1}^{q_i}$ denote a vector of characteristics associated with a $q_i \times 1$ random effect $\mathbf{b}_i$ for the j^{th} element within the i^{th} cluster, where vec is an operator such that $vec\{\bullet_k\}_{k=1}^{K} = (\bullet_1^T, \ldots, \bullet_K^T)^T$. Let $\mathbf{x}_{ij} = vec\{x_{kij}\}_{k=1}^{p}$, with $x_{1ij} = 1$, denote the vector of characteristics associated with fixed effect $\boldsymbol{\beta}$ for the j^{th} element within the i^{th} cluster. Let ϵ_{ij} be the random disturbance and the linear mixed model we considered can be expressed as

$$Y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i + \epsilon_{ij} \qquad (2.1)$$

for $i = 1, \ldots, M, j = 1, \ldots, N_i$.

Model (2.1) can be expressed in a matrix form. Let $\mathbf{Y}_i = vec\{Y_{ij}\}_{j=1}^{N_i}$ and $\mathbf{Y} = vec\{\mathbf{Y}_i\}_{i=1}^{M}$, a $N \times 1$ vector, where $N = \sum_{i=1}^{M} N_i$. Similarly, let $\boldsymbol{\epsilon}_i = vec\{\epsilon_{ij}\}_{j=1}^{N_i}$, and $\boldsymbol{\epsilon} = vec\{\boldsymbol{\epsilon}_i\}_{i=1}^{M}$. Denote $vec\{\mathbf{b}_i\}_{i=1}^{M}$ as $\mathbf{b}$. Let $\mathbf{Z}_i = vec\{\mathbf{z}_{ij}^T\}_{j=1}^{N_i}$ and $\mathbf{Z} = \bigoplus\{\mathbf{Z}_i\}_{i=1}^{M}$, where $\bigoplus$ denotes the direct sum. Matrix $\mathbf{Z}$ is block diagonal with $\mathbf{Z}_i, i = 1, \ldots, M$ as its diagonal blocks. Let $\mathbf{X}_i = vec\{\mathbf{x}_{ij}^T\}_{j=1}^{N_i}$ and $\mathbf{X} = vec\{\mathbf{X}_i\}_{i=1}^{M}$. The matrix expression of (2.1) is then:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\,\mathbf{b} + \boldsymbol{\epsilon} \qquad (2.2)$$

The random effects $\mathbf{b}_i$ and the disturbance $\boldsymbol{\epsilon}_i$ are assumed to have mean $\mathbf{0}$ and variance-covariance matrices $\mathbf{D}_i$ and $\mathbf{R}_i$, respectively. We assume that $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ are independent. The variance-covariance matrix $\mathbf{R} = cov(\mathbf{Y}|\mathbf{b})$ is normally made up of diagonal blocks, $\mathbf{R} = \bigoplus\{\mathbf{R}_i\}_{i=1}^{M}$. The random effects $\mathbf{b}_i, i = 1, \ldots, M$ are usually independent from one another, which implies that the variance-covariance matrix of $\mathbf{b}$ is $\mathbf{D} = \bigoplus\{\mathbf{D}_i\}_{i=1}^{M}$. The resulting variance-covariance matrix of responses $\mathbf{Y}$ is $\mathbf{V} = \mathbf{R} + \mathbf{Z}\mathbf{D}\mathbf{Z}^T$, where $\mathbf{V} = \bigoplus\{\mathbf{V}_i\}_{i=1}^{M}$ with blocks $\mathbf{V}_i = \mathbf{R}_i + \mathbf{Z}_i\mathbf{D}_i\mathbf{Z}_i^T, i = 1, \ldots, M$.

2.1 Nested error linear regression model

The nested error linear regression model is a special case of model (2.1). Battese, Harter and Fuller (1988), Prasad and Rao (1990), etc., used the nested error linear regression model to estimate the mean of small areas. In the nested error linear regression model, random effects in the model are essentially the area effects. Thus, $\mathbf{b}_i$ becomes a scalar denoted as b_i with covariates $\mathbf{z}_{ij} = 1$. For each observation Y_{ij}, we have

$$Y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + b_i + \epsilon_{ij} \qquad (2.3)$$

Random area effects $b_i, i = 1, \ldots, M$ are independent to one another with mean 0 and a common variance σ_b^2, i.e., $\mathbf{D} = cov(\mathbf{b}) = \sigma_b^2 \mathbf{I}_M$. Random errors $\epsilon_{ij}, i = 1, \ldots, M, j = 1, \ldots, N_i$ are assumed to be independent with a mean 0 and a common variance σ_e^2, i.e., $\mathbf{R} = cov(\boldsymbol{\epsilon}) = \sigma_e^2 \mathbf{I}_N$.

2.2 Simple random effects model

In a simple random effects model, all elements share a common mean, i.e., the fixed effects $\boldsymbol{\beta}$ reduces to a scalar μ. Consequently, covariate $\mathbf{x}_{ij} = 1$. Random effects are essentially the cluster effects. Vector $\mathbf{b}_i$ becomes a scalar denoted as b_i with covariates $\mathbf{z}_{ij} = 1$, which are the same as in the nested error linear regression model (2.3). The simple random effects model can be expressed as

$$Y_{ij} = \mu + b_i + \epsilon_{ij}, \qquad (2.4)$$

which is also known as a one-way classification model in ANOVA.

3 Estimation

3.1 Sampling design set-up

Given the population model (2.2), we can obtain a corresponding census estimating equation (CEE). Since population data are usually not available in practice, we estimate parameters in the CEE using a sample. We assume a two-stage sampling design. At stage 1, m level 2 clusters are selected from the M population clusters with inclusion probability $\pi_i, i = 1, \ldots, m$. At stage 2, n_i elements are selected with inclusion probability $\pi_{j|i}$ from the N_i elements within the i^{th} selected cluster. Totally, $n = \sum_i n_i$ units are sampled from the population. The weight associated with the i^{th} cluster is $w_i = \pi_i^{-1}$, whereas the one associated with the j^{th} selected element in the i^{th} selected cluster is $w_{j|i} = \pi_{j|i}^{-1}$. Hence, the inclusion probability associated with the ij^{th} unit is $\pi_{ij} = \pi_i \pi_{j|i}$ and its weight is $w_{ij} = w_i w_{j|i}$. Let $\mathbf{W}_{.|i} = \bigoplus\{w_{j|i}\}_{j=1}^{n_i}$ and $\mathbf{W} = \bigoplus\{w_i \mathbf{W}_{.|i}\}_{i=1}^{m}$ be the matrices of second stage weights and final weights, respectively.

We further assume that the sample data $\mathbf{y}$ follow the same model as the population data,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}. \qquad (3.1)$$

The random effects $\mathbf{b}$ and the disturbance $\mathbf{e}$ are assumed to have mean $\mathbf{0}$ and variance-covariance matrices $\mathbf{D}$ and $\mathbf{R}$, respectively. The variance-covariance matrix $\mathbf{R} = cov(\mathbf{y}|\mathbf{b}) = \bigoplus\{\mathbf{R}_i\}_{i=1}^m$, where $\mathbf{R}_i$ is a sub-matrix of $\mathbf{R}_i = cov(\mathbf{Y}_i|\mathbf{b}_i)$, $\mathbf{D} = \bigoplus\{\mathbf{D}_i\}_{i=1}^m$. The resulting variance-covariance matrix of responses $\mathbf{y}$ is $\mathbf{V} = \mathbf{R} + \mathbf{Z}\mathbf{D}\mathbf{Z}^T$, where $\mathbf{V} = \bigoplus\{\mathbf{V}_i\}_{i=1}^m$ with blocks $\mathbf{V}_i = \mathbf{R}_i + \mathbf{Z}_i\mathbf{D}_i\mathbf{Z}_i^T, i = 1, \ldots, m$.

Under model (3.1), the unweighted sample estimating equation (UWSEE) $\mathbf{\Phi}_0$ is

$$\mathbf{\Phi}_0(\boldsymbol{\beta}, \mathbf{b}) = \begin{bmatrix} \mathbf{\Phi}_{01} \\ \mathbf{\Phi}_{02} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{\Phi}_{03} \end{bmatrix} = \mathbf{0} \qquad (3.2)$$

where $\mathbf{\Phi}_{01}(\mathbf{y}, \boldsymbol{\beta}, \mathbf{b}) = \mathbf{X}^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})$, $\mathbf{\Phi}_{02}(\mathbf{y}, \boldsymbol{\beta}, \mathbf{b}) = vec\{\mathbf{\Phi}_{02i}(\mathbf{y}_i, \boldsymbol{\beta}, \mathbf{b}_i)\}_{i=1}^m$, $\mathbf{\Phi}_{02i}(\mathbf{y}_i, \boldsymbol{\beta}, \mathbf{b}_i) = \mathbf{Z}_i^T\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)$ and $\mathbf{\Phi}_{03}(\mathbf{b}) = vec\{\mathbf{\Phi}_{03i}(\mathbf{b}_i)\}_{i=1}^m$, and $\mathbf{\Phi}_{03i}(\mathbf{b}_i) = -\mathbf{D}_i^{-1}\mathbf{b}_i$ for $i = 1, \ldots, m$.

Investigations (e.g. Pfeffermann, et al. (1998)) showed that estimates obtained by solving the UWSEE can be biased when sampling is informative. Incorporating design weights, we construct a WSEE as

$$\mathbf{\Phi}^*(\boldsymbol{\beta}, \mathbf{b}) = \begin{bmatrix} \hat{\mathbf{\Phi}}_1 \\ \hat{\mathbf{\Phi}}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{\Omega}^{-1}\mathbf{\Phi}_3 \end{bmatrix} = \mathbf{0} \qquad (3.3)$$

where $\hat{\mathbf{\Phi}}_1 = \mathbf{X}^T\mathbf{R}^{-1}\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})$, $\hat{\mathbf{\Phi}}_{2i} = \mathbf{Z}_i^T\mathbf{R}_i^{-1}\mathbf{W}_{.|i}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i), i = 1 \ldots, m$. Define $\mathbf{\Omega} = \bigoplus\{\mathbf{\Omega}_i\}_{i=1}^m$ as a matrix of the first order inclusion probabilities π_i, $i = 1, \ldots, m$, with $\mathbf{\Omega}_i = \pi_i\mathbf{I}_{q_i}$. Let $\hat{\mathbf{\Phi}}_2 = vec\{\pi_i^{-1}\hat{\mathbf{\Phi}}_{2i}\}_{i=1}^m = \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b})$. It can be shown that when the variance-covariance matrix $cov(\mathbf{Y}|b) = \boldsymbol{R}$ is diagonal, $\hat{\mathbf{\Phi}}_1$ is the optimal estimator for $\mathbf{\Phi}_1$ in the CEE by Theorem 1 in Godambe and Thompson (1986). A similar argument fits for $\hat{\mathbf{\Phi}}_{2i}$.

Solving the WSEE (3.2), we obtain the pseudo BLUP for $\boldsymbol{\beta}$ and $\mathbf{b}$ as

$$\hat{\boldsymbol{\beta}} = \{\mathbf{X}^T[\mathbf{V}^*]^{-1}\mathbf{X}\}^{-}\mathbf{X}^T[\mathbf{V}^*]^{-1}\mathbf{y} \qquad (3.4)$$

and

$$\hat{\mathbf{b}}_i = \mathbf{\Omega}_i\mathbf{D}_i\mathbf{Z}_i^T[\mathbf{V}_i^*]^{-1}(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}) \qquad (3.5)$$

where $\mathbf{V}^* = \bigoplus\{\mathbf{V}_i^*\}_{i=1}^m$ with $\mathbf{V}_i^* = \pi_i[\mathbf{W}_{.|i}^{-1}\mathbf{R}_i + \mathbf{Z}_i\mathbf{D}_i\mathbf{Z}_i^T]$.

When the variance components are known, the covariances of $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{b}}_i$ are

$$\begin{aligned} cov(\hat{\boldsymbol{\beta}}) &= \{\mathbf{X}^T[\mathbf{V}^*]^{-1}\mathbf{X}\}^{-}\{\mathbf{X}^T[\mathbf{V}^*]^{-1}\mathbf{V} \\ &\quad [\mathbf{V}^*]^{-1}\mathbf{X}\}\{\mathbf{X}^T[\mathbf{V}^*]^{-1}\mathbf{X}\}^{-} \end{aligned} \qquad (3.6)$$

and

$$\begin{aligned} cov(\hat{\mathbf{b}}_i) = \{&\mathbf{\Omega}_i\mathbf{D}_i\mathbf{Z}_i^T[\mathbf{V}_i^*]^{-1}\}\mathbf{C} \\ &\{\mathbf{\Omega}_i\mathbf{D}_i\mathbf{Z}_i^T[\mathbf{V}_i^*]^{-1}\}^T \end{aligned} \qquad (3.7)$$

where $\mathbf{C} = cov(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}) = \mathbf{V}_i + \mathbf{X}_i cov(\hat{\boldsymbol{\beta}})\mathbf{X}_i^T - 2\mathbf{V}_i[\mathbf{V}_i^*]^{-1}\mathbf{X}_i[\mathbf{X}^T[\mathbf{V}^*]^{-1}\mathbf{X}]^{-}\mathbf{X}_i^T$.

The function $\hat{\mathbf{\Phi}}_{2i} + \hat{\mathbf{\Phi}}_{3i}$ in WSEE (3.3) does not reduce to $\mathbf{\Phi}_{02i} + \mathbf{\Phi}_{03i}$ in simple SEE (3.2) for simple random sampling (SRS) design. To adjust for SRS design, we define $\mathbf{\Omega}_i^\dagger = \frac{\sum_j w_{ij}}{\sum_j w_{ij}^2}\mathbf{I}_{q_i}$ and $\mathbf{\Omega}^\dagger = \bigoplus\{\mathbf{\Omega}_i^\dagger\}_{i=1}^m$. The WSEE adjusted for SRS is $\mathbf{\Phi}^\dagger$

$$\mathbf{\Phi}^\dagger(\boldsymbol{\beta}, \mathbf{b}) = \begin{bmatrix} \hat{\mathbf{\Phi}}_1 \\ \hat{\mathbf{\Phi}}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ [\mathbf{\Omega}^\dagger]^{-1}\mathbf{\Phi}_3 \end{bmatrix} = \mathbf{0}. \quad (3.8)$$

In the case of SRS at cluster level and element level, $\mathbf{W}_{.|i} = \frac{N_i}{n_i}\mathbf{I}_{n_i}$ and $\mathbf{W} = \frac{M}{m}\bigoplus\{\frac{N_i}{n_i}\mathbf{I}_{n_i}\}_{i=1}^m$. The matrix $\mathbf{\Omega}_i^\dagger = \frac{M}{m}\frac{N_i}{n_i}\mathbf{I}_{n_i}$. The i^{th} component of equation $\hat{\mathbf{\Phi}}_2 + [\mathbf{\Omega}^\dagger]^{-1}\mathbf{\Phi}_3 = 0$ in (3.7) reduces $\mathbf{\Phi}_{02i} + \mathbf{\Phi}_{03i} = \mathbf{0}$ as in the simple SEE (3.2).

Solving $\mathbf{\Phi}^\dagger$, we obtain another set of pseudo BLUP $\hat{\boldsymbol{\beta}}^\dagger$, $\hat{\mathbf{b}}_i^\dagger$ and their variance-covariance matrices by substituting $\mathbf{V}^*$ by $\mathbf{V}^\dagger$ and $\mathbf{\Omega}$ by $\mathbf{\Omega}^\dagger$ in 3.4-3.7, where $\mathbf{V}^\dagger = \mathbf{W}^{-1}\mathbf{R} + \mathbf{\Omega}^\dagger\mathbf{Z}\mathbf{D}\mathbf{Z}^T$, $\mathbf{C}^\dagger = cov(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}) = \mathbf{V}_i + \mathbf{X}_i cov(\hat{\boldsymbol{\beta}})\mathbf{X}_i^T - 2\mathbf{V}_i[\mathbf{V}_i^\dagger]^{-1}\mathbf{X}_i[\mathbf{X}^T[\mathbf{V}^\dagger]^{-1}\mathbf{X}]^{-}\mathbf{X}_i^T$.

When the variance components are unknown, we obtain the pseudo EBLUP for $\boldsymbol{\beta}$ and $\mathbf{b}$ and their covariances by plugging in the consistent estimates of variance components as discussed in section 4.

3.2 Nested error linear regression model

The variance-covariance matrix for the random effects vector $\mathbf{b}$ is $\mathbf{D} = \sigma_b^2 \mathbf{I}_m$, for the random errors is $\mathbf{R} = \sigma_e^2 \mathbf{I}_n$, and the covariate matrix is $\mathbf{Z}_i = \mathbf{1}_{n_i}$. Consequently, the covariance of $\mathbf{y}_i$ is $\mathbf{V}_i = \sigma_b^2 \mathbf{J}_{n_i} + \sigma_e^2 \mathbf{I}_{n_i}$, where $\mathbf{J}_{n_i} = \mathbf{1}_{n_i}\mathbf{1}_{n_i}^T$. Using (3.4) and (3.5), we obtain the pseudo BLUP estimator for fixed effects as

$$\hat{\boldsymbol{\beta}} = [\sum_{ij} \mathbf{x}_{ij}\mathbf{d}_{ij}^T]^{-1}[\sum_{ij} y_{ij}\mathbf{d}_{ij}] \qquad (3.9)$$

and that for random effects b_i as

$$\hat{b}_i = \gamma_i(\overline{y}_{iw} - \overline{\mathbf{x}}_{iw}^T\hat{\boldsymbol{\beta}}) \qquad (3.10)$$

where $\mathbf{d}_{ij} = w_{ij}(\mathbf{x}_{ij} - \gamma_i\overline{\mathbf{x}}_{iw})$ and $\gamma_i = \sigma_b^2/(\sigma_e^2/\sum_j w_{j|i} + \sigma_b^2)$, $\overline{y}_{iw} = \sum_j w_{j|i}y_{ij}/\sum_j w_{j|i}$ and $\overline{\mathbf{x}}_{iw} = \sum_j w_{j|i}\mathbf{x}_{ij}/\sum_j w_{j|i}$. The estimators (3.9) agrees with that obtained by Pfeffermann, et al. (1998). It is straightforward to obtain the corresponding variance-covariance matrices of $\hat{\boldsymbol{\beta}}$ and $\hat{b}_i, i = 1, \ldots, m$.

The pseudo BLUP adjusted for SRS can be obtained by solving the WSEE (3.8). Resulting estimators $\hat{\boldsymbol{\beta}}^\dagger$ and $\hat{b}_i^\dagger$ have the same expression as (3.9) and (3.10), respectively, with γ_i replaced by $\gamma_i^\dagger = \sigma_b^2/(\sigma_e^2 \frac{\sum_j w_{ij}^2}{(\sum_j w_{ij})^2} + \sigma_b^2)$. Estimators $\hat{\boldsymbol{\beta}}^\dagger$ and $\hat{b}_i^\dagger$ we obtained for the nested error linear regression model are in coincidence with those by You and Rao (2002).

3.3 Simple random effects model

Substituting $x_{ij} = 1$ into (3.9) and (3.10), we obtain the pseudo BLUP estimator of the mean as

$$\hat{\mu} = \frac{\sum_{ij} w_{ij}(1-\gamma_i)\overline{y}_{iw}}{\sum_{ij} w_{ij}(1-\gamma_i)} \qquad (3.11)$$

and that for the random effects

$$b_i = \gamma_i(\overline{y}_{iw} - \hat{\mu}). \qquad (3.12)$$

The respective variances of $\hat{\mu}$ and $\hat{b}_i$ can be obtained when σ_e^2 and σ_b^2 are known.

Replacing γ_i by $\gamma_i^\dagger$ in (3.11) and (3.12) results in the pseudo BLUP $\hat{\mu}^\dagger$ and $\hat{b}_i^\dagger$, respectively. Note that estimator $\hat{\mu}^\dagger$ reduces to $\hat{\mu}_0 = \frac{\sum_i n_i(1-\tilde{\gamma}_i)\overline{y}_i}{\sum_i n_i(1-\tilde{\gamma}_i)}$ in the case of SRS, where $\tilde{\gamma}_i = \sigma_b^2/(\sigma_e^2/n_i + \sigma_b^2)$. The estimator $\hat{\mu}_0$ coincides with the one obtained by applying the Prasad and Rao (1990) procedure on the simple random effects model. Note that the method of Prasad and Rao (1990) does not account for sampling weights. Estimators $\hat{\mu}$ and $\hat{\mu}^\dagger$ result from using both individual level and aggregated level models. Prasad and Rao (1999) proposed an additional estimator $\hat{\mu}^* = \sum_i \gamma_i^\dagger \overline{y}_{iw}/\sum_i \gamma_i^\dagger$ for μ using only an aggregated level model. We compare these four estimators for μ ($\hat{\mu}$, $\hat{\mu}^\dagger$, $\hat{\mu}_0$, $\hat{\mu}^*$) with a simulation study in Section 6.

4 Estimation of variance components

Maximum likelihood (ML) method, restricted maximum likelihood (REML) method and the EM algorithm have been used for estimating the variance-covariance of response vector $\mathbf{y}$. Pfeffermann et al. (1998) proposed an iterative process for estimating the variance-covariance matrix in model (2.2) using only first-order selection probabilities.

For the special case of the nested error linear regression model, Prasad and Rao (1990) and You and Rao (2002) adopted Henderson's method III to estimate the variance components σ_b^2 and σ_e^2. However, the accuracy of these estimated variances is not known in the design based context. Furthermore, they do not take into account the sampling design.

One way to account for the sampling design is to duplicate the sample data using sample weights to obtain a pseudo population. The resulting pseudo population resembles the true population in so far as that the sampled cluster i can be considered to have been selected with probability π_i and that the sampled element j within cluster i can be considered to have been selected with probability $\pi_{j|i}$ from this pseudo population.

The duplication occurs in two steps. We

assume that sampling weights $\{w_i\}$'s and $\{w_{j|i}\}$'s are integers. First, the response y_{ij} and the covariates $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$ observed on the ij^{th} element are duplicated $w_{j|i}$ times. The inflated response vector is $\tilde{\mathbf{y}}_i = vec\{\mathbf{1}_{w_{j|i}} \bigotimes y_{ij}\}_{j=1}^{n_i}$. Similarly, the inflated covariate matrix is $\tilde{\mathbf{X}}_i = vec\{\mathbf{1}_{w_{j|i}} \bigotimes \mathbf{x}_{ij}^T\}_{j=1}^{n_i}$ and $\tilde{\mathbf{Z}}_i = vec\{\mathbf{1}_{w_{j|i}} \bigotimes \mathbf{z}_{ij}^T\}_{j=1}^{n_i}$.

Second, the inflated vector $\tilde{\mathbf{y}}_i$, the covariate matrices $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{Z}}_i$ are duplicated w_i times to form the response vector $\tilde{\mathbf{y}}$, covariate matrices $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Z}}$ of the pseudo population. This results in $\tilde{\mathbf{y}} = vec\{\mathbf{1}_{w_i} \bigotimes \tilde{y}_i\}_{i=1}^m$, $\tilde{\mathbf{X}} = vec\{\mathbf{1}_{w_i} \bigotimes \tilde{X}_i\}_{i=1}^m$, and $\tilde{\mathbf{Z}} = \bigoplus\{\mathbf{I}_{w_i} \bigotimes \tilde{\mathbf{Z}}_i\}_{i=1}^m$.

Denote inflated random disturbances as $\tilde{\boldsymbol{\epsilon}}$ and inflated random effects as $\tilde{\mathbf{b}}$. Their respective variance-covariance matrices are $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{D}}$. Here, we assume that the components of $\tilde{\boldsymbol{\epsilon}}$, $\{\tilde{\epsilon}_{ij}\}$'s, are independent. Also, the components of $\tilde{\mathbf{b}}$, $\{\tilde{b}_i\}$'s, are independent.

The resulting model for this pseudo population is

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\mathbf{Z}}\tilde{\mathbf{b}} + \tilde{\boldsymbol{\epsilon}} \qquad (4.1)$$

and the variance-covariance matrix of $\tilde{\mathbf{y}}$ is $\tilde{\mathbf{V}} = \tilde{\mathbf{R}} + \tilde{\mathbf{Z}}\tilde{\mathbf{D}}\tilde{\mathbf{Z}}^T$.

Using Henderson's method III on the pseudo population, the sum of square errors (SSE) for fitting a nested error linear regression model (4.1) on the pseudo population is $SSE = \sum_{ij} w_{ij}(y_{ij} - \overline{y}_{iw})^2 - \sum_{ij} w_{ij}(y_{ij} - \overline{y}_{iw})(\mathbf{x}_{ij} - \overline{\mathbf{x}}_{iw})^T[\sum_{ij} w_{ij}(\mathbf{x}_{ij} - \overline{\mathbf{x}}_{iw})(\mathbf{x}_{ij} - \overline{\mathbf{x}}_{iw})^T]^{-1} \sum_{ij} w_{ij}(y_{ij} - \overline{y}_{iw})(\mathbf{x}_{ij} - \overline{\mathbf{x}}_{iw})$ and the estimator for σ_e^2 is

$$\hat{\sigma}_{ew}^2 = SSE/(\sum_{ij} w_{ij} - \sum_i w_i - p + 1). \qquad (4.2)$$

Denote the difference between the reduction of sum of squares due to fitting model (4.1) and that due to fitting model $\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}}$ as $S(\mathbf{b}|\boldsymbol{\beta})$, where $S(\mathbf{b}|\boldsymbol{\beta}) = \sum_{ij} w_{ij} y_{ij}^2 - (\sum_{ij} w_{ij} y_{ij} \mathbf{x}_{ij}^T)(\sum_{ij} w_{ij}\mathbf{x}_{ij}\mathbf{x}_{ij}^T)^{-1}(\sum_{ij} w_{ij} y_{ij}\mathbf{x}_{ij}) - SSE$. Consequently, the variance component σ_b^2 can be estimated by

$$\hat{\sigma}_{bw}^2 = [S(\mathbf{b}|\boldsymbol{\beta}) - (\sum_i w_i - 1)\hat{\sigma}_{ew}^2]/(\sum_{ij} w_{ij} - t_1), \qquad (4.3)$$

where $t_1 = tr[(\sum_{ij} w_{ij}\mathbf{x}_{ij}\mathbf{x}_{ij}^T)^{-1} \sum_i w_i (\sum_j w_{j|i})^2 \overline{\mathbf{x}}_{iw}\overline{\mathbf{x}}_{iw}^T]$.

For the simple random effects model, by substituting $\mathbf{x}_{ij} = 1$ and $p = 1$ into (4.2) and (4.3), the estimators for variance components are obtained as

$$\hat{\sigma}_{e0w}^2 = \sum_{ij} w_{ij}(y_{ij} - \overline{y}_{iw})^2/(\sum_{ij} w_{ij} - \sum_i w_i) \qquad (4.4)$$

and

$$\hat{\sigma}_{b0w}^2 = \frac{\sum_{ij} w_{ij}(\overline{y}_{iw} - \overline{y}_w)^2 - (\sum_i w_i - 1)\hat{\sigma}_e^2}{\sum_{ij} w_{ij} - \sum_i w_i(\sum_j w_{j|i})^2/\sum_{ij} w_{ij}} \qquad (4.5)$$

where $\overline{y}_w = \sum_{ij} w_{ij}y_{ij}/\sum_{ij} w_{ij}$.

Estimators (4.4) and (4.5) are identical to those obtained by Graubard and Korn (1996) ($\hat{\sigma}_{eC}^2$ and $\hat{\sigma}_{aC}^2$), respectively. As noted by Graubard and Korn, estimators (4.4) and (4.5) do not reduce to simple sample estimators $\hat{\sigma}_{e0u}^2 = \sum_{ij}(y_{ij} - \overline{y}_i)^2/\sum_i(n_i - 1)$ and $\hat{\sigma}_{b0u}^2 = [\sum_{ij}(y_{ij} - \overline{y})^2 - (n-1)\hat{\sigma}_e^2]/(n - \sum_i n_i^2/n)$ (denoted as $\hat{\sigma}_{eA}^2$ and $\hat{\sigma}_{bA}^2$ in Graubard and Korn (1996)) for some non-informative sample designs, where $\overline{y}_i = \sum_j y_{ij}/n_i$.

Furthermore, under model (2.3) for sample data, estimators (4.2) and (4.3), and thus (4.4) and (4.5) are not model unbiased. To adjust for the bias, we calculate the expectation of SSE and $S(\mathbf{b}|\boldsymbol{\beta})$ under model (2.3). Unbiased estimators for σ_e^2 and σ_b^2 are obtained as

$$\hat{\sigma}_{e_{adj}}^2 = SSE/[\sum_{ij} w_{ij} - \sum_i \frac{\sum_j w_{ij}^2}{\sum_j w_{ij}} - t_2)] \qquad (4.6)$$

and

$$\hat{\sigma}_{b_{adj}}^2 = [\sum_{ij} w_{ij}y_{ij}^2 - (\sum_{ij} w_{ij}y_{ij}\mathbf{x}_{ij}^T)(\sum_{ij} w_{ij}\mathbf{x}_{ij}\mathbf{x}_{ij}^T)^{-1} \sum_{ij} w_{ij}y_{ij}\mathbf{x}_{ij}) - (\sum_{ij} w_{ij} - t_3)\hat{\sigma}_e^2]/(\sum_{ij} w_{ij} - t_4), \qquad (4.7)$$

where $t_2 = tr\{[\sum_{ij} w_{ij}(\mathbf{x}_{ij} - \overline{\mathbf{x}}_{iw})(\mathbf{x}_{ij} - \overline{\mathbf{x}}_{iw})^T]^- \sum_{ij} w_{ij}^2(\mathbf{x}_{ij} - \overline{\mathbf{x}}_{iw})(\mathbf{x}_{ij} - \overline{\mathbf{x}}_{iw})^T\}$, $t_3 = tr\{[\sum_{ij} w_{ij}\mathbf{x}_{ij}\mathbf{x}_{ij}^T]^- \sum_{ij} w_{ij}^2\mathbf{x}_{ij}\mathbf{x}_{ij}^T\}$, and $t_4 = tr\{\sum_{ij} w_{ij}\mathbf{x}_{ij}\mathbf{x}_{ij}^T]^{-1} \sum_i(\sum_j w_{ij})^2 \overline{\mathbf{x}}_{iw}\overline{\mathbf{x}}_{iw}^T\}$.

Estimators (4.6) and (4.7) reduce to simple sample estimators resulting from Henderson's method III that were used by Prasad and Rao (1990) and You and Rao (2002) for some non-informative sampling designs.

For the simple random effects model (2.4), the model unbiased estimators for σ_e^2 and σ_b^2 are obtained by substituting $x_{ij} = 1$ into (4.6) and (4.7)

$$\hat{\sigma}^2_{e0_{adj}} = \frac{\sum_{ij} w_{ij}(y_{ij} - \overline{y}_{iw})^2}{\sum_{ij} w_{ij} - \sum_i \frac{\sum_j w_{ij}^2}{\sum_j w_{ij}}} \qquad (4.8)$$

and

$$\hat{\sigma}^2_{b0_{adj}} = \frac{\sum_{ij} w_{ij}(\overline{y}_{iw} - \overline{y}_w)^2 - c\hat{\sigma}_e^2}{\sum_{ij} w_{ij} - \frac{\sum_i(\sum_j w_{ij})^2}{\sum_{ij} w_{ij}}}, \qquad (4.9)$$

where $c = \sum_i \sum_j w_{ij}^2 / \sum_j w_{ij} - \sum_{ij} w_{ij}^2 / \sum_{ij} w_{ij}$.

Estimators $\hat{\sigma}^2_{e0_{adj}}$ and $\hat{\sigma}^2_{b0_{adj}}$ reduce to $\hat{\sigma}^2_{e0u}$ and $\hat{\sigma}^2_{b0u}$ for some non-informative sample designs considered by Graubard and Korn (1996) (e.g. SRS with $w_i = M/m$ and $w_{j|i} = N/n$).

5 Model and Design-based properties of variance components

In this section, we will investigate the model and design based properties of variance components for the simple random effects model (2.4) as well as others discussed in Korn and Graubard (2003).

Given that the population values Y_{ij} are assumed to satisfy the simple random effects model (2.4), we are interested in estimating the variances of the $\{b_i\}$'s and $\{\epsilon_{ij}\}$'s, namely σ_b^2 and σ_e^2. The model-based variance components S_e^2 and S_b^2 are estimated via the methods of moments (Searle et al. (1992), page 106). Let

$$S_e^2 = \frac{1}{\sum_{i=1}^M (N_i - 1)} \sum_{i=1}^M \sum_{j=1}^{N_i} (Y_{ij} - \bar{Y}_i)^2 \qquad (5.1)$$

and

$$S_b^2 = \frac{1}{N_0}(M-1)\sum_{i=1}^M N_i(\bar{Y}_i - \bar{Y})^2 - \frac{S_e^2}{N_0} \qquad (5.2)$$

where $N_0 = \frac{1}{M-1}(\sum_{i=1}^M N_i - \frac{\sum_{i=1}^M N_i^2}{\sum_{i=1}^M N_i})$, $\overline{Y}_i$ is the mean of the population observations in the i^{th} cluster and $\overline{Y}$ is the overall mean. The two-stage sampling design given in section 3.2 is used to draw the sample. As noted in section 4, estimators (4.4) and (4.5) are identical to the $\hat{\sigma}^2_{ec}$ and $\hat{\sigma}^2_{bc}$ in Graubard and Korn (1996).

We provide the design-based and model-based properties of the following variance component estimators given in section 4, as well as those provided in Graubard and Korn (1996), and Korn and Graubard (2003).

The various variance components estimators for σ_e^2 are

$$\hat{\sigma}^2_{eA} = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i)^2}{\sum_{i=1}^m (n_i - 1)}, \qquad (5.3)$$

$$\hat{\sigma}^2_{eB} = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} w_{ij}(y_{ij} - \overline{y}_{iw})^2}{\sum_{i=1}^m \sum_{j=1}^{n_i} (w_{ij} - 1)}, \qquad (5.4)$$

$$\hat{\sigma}^2_{eC} = \frac{\sum_{i=1}^m w_i \sum_{j=1}^{n_i} w_{j|i}(y_{ij} - \overline{y}_{iw})^2}{\sum_{i=1}^m w_i(\sum_{j=1}^{n_i} w_{j|i} - 1)}, and \qquad (5.5)$$

$$\hat{\sigma}^2_{eD} = \frac{\sum_{i=1}^m \overline{w}_i \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i)^2}{\sum_{i=1}^m \overline{w}_i(n_i - 1)}, \qquad (5.6)$$

where $\overline{w}_i = \sum_{j=1}^{n_i} w_{ij}/n_i$ is the within cluster i sample mean of the weights w_{ij}, and

$$\hat{\sigma}^2_{eKG} = \frac{\frac{1}{2}\sum_{i=1}^m \frac{(N_i-1)I(n_i>1)}{\pi_i} \frac{\sum_{k<j}^{n_i} \frac{(y_{ik}-y_{ij})^2}{\pi^{(2)}_{(kj)|i}}}{\sum_{k<j}^{n_i} \frac{1}{\pi^{(2)}_{(kj)|i}}}}{\sum_{i=1}^m \frac{(N_i-1)I(n_i>1)}{\pi_i}} \qquad (5.7)$$

where $\pi^{(2)}_{(kj)|i}$ denotes the conditional joint inclusion probability for units k and j within the i^{th} sampled cluster.

Graubard and Korn (1996) reported on the various weaknesses (biases) of estimators (5.3) - (5.6), and favored $\hat{\sigma}^2_{eD}$ (5.6). Estimator (5.7), proposed by Korn and Graubard (2003) is approximately design unbiased. The design unbiasedness, however, requires the computation of the second-stage joint inclusion probabilities $\pi^{(2)}_{(kj)|i}$ for units k and j within each sampled cluster i.

It should be noted that an alternative estimator to S_e^2 that bypasses the estimation of $\frac{N_i(N_i-1)}{2}$ is:

$$\hat{\sigma}^2_{eHH} = \frac{\sum_{i=1}^m \frac{1}{\pi_i N_i} \sum_{k<j} \frac{(y_{ik}-y_{ij})^2}{\pi^{(2)}_{(kj)|i}}}{\sum_{i=1}^m \frac{N_i-1}{\pi_i}}. \qquad (5.8)$$

This estimator should be less variable than $\hat{\sigma}^2_{eKG}$ because it uses more information (namely the incorporation of the $\{N_i\}$'s).

The design-based expectation $(E_p(\hat{\theta}))$ of the above sample estimators of σ^2_e then follows. Denote the universe of clusters as U, the first-stage sample of clusters as s_1, the set of elements within i^{th} sampled cluster as s_{1i} $(i = 1, ..., m)$, and the set of elements sampled within the elements of the i^{th} sampled cluster s_{1i} as s_{2i}.

The design based expectation of $\hat{\sigma}^2_{eA}$ is given by $E_p(\hat{\sigma}^2_{eA}) = E_{p1}E_{p2}(\hat{\sigma}^2_{eA}|s_1)$, where $E_{pi}(i = 1, 2)$ denotes the design expectation for each sampling stage. It can be shown that $E_p(\hat{\sigma}^2_{eA}) = \{\sum_{i=1}^{M} \frac{\pi_i}{n_i}[(n_i - 1)\sum_{j=1}^{N_i} \pi_{j|i} y^2_{ij} - 2\sum\sum_{j<k}^{N_i} \pi^{(2)}_{jk|i} y_{ij} y_{ik}]\} / \sum_{i=1}^{M}(n_i - 1)\pi_{ij} \neq S^2_E$. This bias still holds in the case of SRS at both stages.

The design-based expectations of $\hat{\sigma}^2_{eB}$, $\hat{\sigma}^2_{eC}$, $\hat{\sigma}^2_{eD}$, $\hat{\sigma}^2_{eKG}$, $\hat{\sigma}^2_{eHH}$, and $\hat{\sigma}^2_{e0_{adj}}$, are obtained using similar derivations.

Model-based expectations are also of interest. For example, using model (2.4) it can be seen that the model-based expectation of $\hat{\sigma}^2_{eA}$ is $\hat{\sigma}^2_{eA} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} E_\xi(y_{ij} - \overline{y}_i)^2}{\sum_{i=1}^{m}(n_i - 1)} = \sigma^2_e$. The design-based and model-based expectations of the variance estimators of S^2_e are summarized in Table 1 for without replacement and arbitrary sampling schemes at both stages.

Table 1: Design and Model based expectations of the various estimators of S^2_e

Estimators	Design-based	Model-based
$\hat{\sigma}^2_{eA}$	Biased	Unbiased
$\hat{\sigma}^2_{eB}$	Biased	Biased
$\hat{\sigma}^2_{eC}$	Biased	Biased
$\hat{\sigma}^2_{eD}$	Biased	Unbiased
$\hat{\sigma}^2_{eKG}$	Unbiased	Unbiased
$\hat{\sigma}^2_{eHH}$	Unbiased	Unbiased
$\hat{\sigma}^2_{e0_{adj}}$	Unbiased	Unbiased

It should be noted that the design-based bias properties of the various estimators of S^2_e, displayed in Table 1, hold for two-stage sampling designs using unequal probability without replacement sampling schemes at each stage. However, there are some exceptions to this. For instance, $\hat{\sigma}^2_{eC}$ is design unbiased whenever $\pi^{(2)}_{(jk)|i} = \pi_{j|i}\pi_{k|i}$, and for arbitrary sampling schemes at the first-stage. Three sampling schemes that satisfy this condition at the second-stage are: (i) simple random sampling without replacement; (ii) pps without replacement; and (iii) Poisson sampling.

6 Simulation

6.1 Procedure

We conducted a simulation study that is similar to the one in Pfeffermann, et al. (1998) to evaluate the performance of the proposed estimators. Estimators for the fixed effects considered in this simulation study are $\hat{\mu}$, $\hat{\mu}^\dagger$, $\hat{\mu}_0$ and $\hat{\mu}^*$ as stated in Section 3. Estimators for variance components we considered are the weighted estimators (4.4) and (4.5), the adjusted estimators (4.8) and (4.9), and the unweighted estimators $(\hat{\sigma}^2_{e0u}, \hat{\sigma}^2_{b0u})$. We generated population values Y_{ij} using a simple random effects model (2.4), where $\mu = 1$, $\sigma^2_e = 1$ and $\sigma^2_b/\sigma^2_e = 0.25$, $i = 1, \ldots, M$ and $j = 1, \ldots, N_i$. We set the number of clusters in the population, M, to 300. Sample cluster sizes studied were $m = 30, 75$. The number of elements in the population, $N_i, i = 1, \ldots, M$ were determined as $exp(r_i)$, where r_i is a random number generated from $N(0, \sigma^2_b)$, truncated by $-1.5\sigma_b$ below and $1.5\sigma_b$ above.

Three sampling schemes were considered: (i) informative at both levels, (ii) informative only at level 2 and (iii) non-informative at both levels. For the informative sampling at both levels, m clusters were selected with probability π_i proportional to some defined 'measure of size' variable S_i, where S_i is defined in the same way as N_i except that random effect b_i was used instead of r_i. The elements were divided into two strata. Stratum 1 included elements whose residue e_{ij} was positive and stratum 2 included the remaining elements. Simple random samples of sizes $.25n_i$ and $.75n_i$ are selected from strata 1 and 2, respectively.

For the informative sampling strictly at level 2, clusters were selected in the same way as in the informative sampling at both levels. Elements were selected by SRSWOR. For non-informative sampling at both levels, clusters were selected using PPS where the measure of size variable S_i was taken as N_i, and elements were selected by SRSWOR.

Five hundred different samples were drawn for each sampling scheme and parameter values. The size of elements, n_i, was either proportional to N_i, $n_i = rN_i$ or fixed as n_0, where n_0 is determined as the average of $rN_i, i = 1, \ldots, M, r = 0.1, 0.4$. For each sample, estimators for fixed effect and variance components are computed.

6.2 Results

Table 2 summarizes the simulation results for estimating the mean and variance components for the various sampling schemes. Estimators that are recommended appear along the rows associated with each sampling scheme. However, all estimators that are enclosed in the shaded area are unbiased, whereas the unshaded ones are biased. These results hold for the two cluster sizes ($m = 30, 75$) that were used in the simulation.

Table 2. Recommended estimators for mean and variance components.

Sampling schemes	Parameters		
	Mean	σ_e^2	σ_b^2
Non-informative	$\hat{\mu}, \hat{\mu}^+, \hat{\mu}_0, \hat{\mu}^*$	$\hat{\sigma}^2_{e0u}, \hat{\sigma}^2_{e0_{adj}}$	$\hat{\sigma}^2_{b0u}, \hat{\sigma}^2_{b0_{adj}}$
Informative-stage 2	$\hat{\mu}, \hat{\mu}^+$	$\hat{\sigma}^2_{e0u}, \hat{\sigma}^2_{e0_{adj}}$	$\hat{\sigma}^2_{b0w}$
Informative-both stages	$\hat{\mu}, \hat{\mu}^+$	$\hat{\sigma}^2_{e0u}, \hat{\sigma}^2_{e0w}$	$\hat{\sigma}^2_{b0w}$

7 Conclusion

In this paper, we proposed a method for estimating fixed effects and random effects involved in a linear mixed model. The proposed estimators are model-unbiased and design consistent. A simulation study shows that the proposed estimators are very competitive for all sampling schema we considered.

For estimating variance components in a nested error regression model, we developed estimators using Henderson's method III incorporating sampling weights. We further derived another set of estimators (4.8) and (4.9) by adjusting the estimators we developed, such that, these estimators would reduce to unweighted estimators for some ignorable sampling design.

A simulation study leads us to recommend using unweighted estimators or adjusted estimators when sampling schema is non-informative. In the case of informative sampling, weighted estimators without adjustment are the best choice.

References

Battese, G.E., Harter, R.M and Fuller, W.A. (1988), "An error-components model for prediction of county crop areas using survey and satellite data," *Journal of the American Statistical Association*, 83, 28-36.

Godambe, V.P. and Thompson, M.E. (1986), "Parameters of superpopulation and survey population: Their relationships and estimation," *International Statistical Review*, 54, 127-138.

Goldstein, H. (1995), *Multilevel Statistical Models*, 2nd edition, London:Arnold.

Graubard, B.I. and Korn, E.L. (1996), "Modeling the sampling design in the analysis of health surveys," *Statistical Methods in Medical Research*, 5, 263-281.

Henderson, C.R. (1975), "Best linear unbiased estimation and prediction under a selection model," *Biometrics*, 31, 423-447.

Korn, E.L. and Graubard, B.I. (2003), "Estimating variance components using survey data," *Journal of the Royal Statistical Society B*, 65, 175.

Laird, N.M. and Ware, J.H. (1982), "Random-effects models for longitudinal data," *Biometrics*, 38, 963-974.

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

Prasad, N.G.N. and Rao, J.N.K. (1990), "The estimation of the mean squared error of small-area estimators," *Journal of the American Statistical Association*, 85, 163-171.

Prasad, N.G.N. and Rao, J.N.K. (1999), "On robust small area estimation using a simple random effects model," *Survey Methodology*, 25, 67-72.

Schall, R. (1991), "Estimation in generalized linear models with random effects," *Biometrika*, 78, 719-727.

Searle, S.R., Casella, G. and McCulloch, C.E. (1992), *Variance Components*. Wiley, New York.

You, Y. and Rao, J.N.K. (2002), "A pseudo empirical best linear unbiased prediction approach to small area estimation using survey weights," *The Canadian Journal of Statistics*, 30, 431-439.