SSC Annual Meeting, May 2010
Proceedings of the Survey Methods Section

A WEIGHTED ESTIMATING EQUATIONS APPROACH TO INFERENCE FOR TWO-LEVEL MODELS FROM SURVEY DATA

J.N.K. Rao[1], F. Verret[2] and M.A. Hidiroglou[3]

ABSTRACT

Multi-level models are extensively used for analyzing survey data with design hierarchy matching the model hierarchy. We propose a weighted estimating equations (WEE) approach for two-level models that leads to design-model consistent estimators of model parameters even when the within-cluster sample sizes are small, provided the number of sample clusters is large. A unified approach based on weighted log composite likelihood that can handle generalized linear multi-level model is also proposed. Results of a small simulation study demonstrate superior performance of the proposed WEE method relative to existing methods under informative sampling within clusters.

KEY WORDS: Small area estimation; unit-level model; informative sampling.

RÉSUMÉ

Les modèles multi niveaux sont souvent utilisés afin d'analyser des données d'enquête lorsque la structure hiérarchique du plan de sondage correspond à la structure hiérarchique du modèle. Nous proposons une approche basée sur des équations d'estimation pondérées (EEP) pour les modèles à deux niveaux menant à des estimateurs qui convergent selon le modèle et selon le plan de sondage aux paramètres du modèle, même si la taille d'échantillon dans les grappes échantillonnées est faible, à condition que le nombre de grappes échantillonné soit suffisamment grand. Une méthode unifiée basée sur la log-vraisemblance composite pondérée pouvant traiter le modèle linéaire généralisé multi niveaux est également proposée. Les résultats d'une petite étude de simulation démontrent que la performance de la méthode par EEP proposée est supérieure aux méthodes existantes si le plan de sondage dans les grappes est informatif.

MOTS CLÉS : Estimation pour les petits domaines; modèle au niveau des unités; plan de sondage informatif.

1. INTRODUCTION

Data collected from large-scale socio-economic, health and other surveys are extensively used for analysis purposes, such as inference on the regression parameters of linear and logistic linear regression population models. Ignoring the survey design features (such as stratification, clustering and unequal selection probabilities) can lead to erroneous inferences on model parameters because of sample selection bias caused by informative sampling. It is tempting to expand the models by including among the auxiliary variables all the design variables that define the selection process at the various levels and then ignore the design and apply standard methods to the expanded model. The main difficulties with this approach are the following (Pfeffermann and Sverchkov, 2003): (1) Not all design variables may be known or accessible to the analyst. (2). Too many design variables can lead to difficulties in making inference from the expanded model. (3) Expanded model may no longer be of scientific interest to the analyst. On the other hand, the design-based approach can provide asymptotically valid repeated sampling inferences without changing the analyst model. A unified approach based on the survey weighted estimating equations leads to design-consistent estimators of the "census" or finite population parameters which in turn estimate the associated model parameters. Further, re-sampling methods, such as the jackknife and the bootstrap for survey data, can provide valid variance estimators and associated inferences on the census parameters. The same methods may also be applicable to inference on the model parameters, in many cases of large-scale surveys. In other cases, it is necessary to estimate the model variance of the census parameters from the sample. The estimator of the total variance is then given by the sum of this estimator and the re-sampling variance estimator. Beaumont and Charest (2010) extended the bootstrap to estimate the total variance associated with the model parameters.

[1] J.N.K. Rao, School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada, K1S 5B6, jrao@math.carleton.ca
[2] F. Verret, Statistics Canada, 16 J, R.-H.-Coats Building, Ottawa, Ontario, Canada, K1A 0T6, Francois.Verret@statcan.gc.ca
[3] M.A. Hidiroglou, Statistics Canada, 16 D, R.-H.-Coats Building, Ottawa, Ontario, Canada, K1A 0T6, Mike.Hidiroglou@statcan.gc.ca

We refer the reader to Rao et al. (2010) for an overview of methods for making inference on regression parameters from complex survey data.

In this paper, our focus is on making design-based inference on the variance component parameters and regression parameters of multi-level models from data obtained from multi-stage sampling designs corresponding to the levels of the model. For example, in an education study of students, schools (first-stage sampling units) may be selected with probabilities proportional to school size and students (second-stage units) within selected schools by stratified random sampling. Again, ignoring the survey design and using traditional methods for multi-level models can lead to erroneous inferences in the presence of sample selection bias. In the design-based approach, estimation of variance component parameters of the model is more difficult than that of regression parameters. Past work on multi-level models for survey data are summarized in Section 2. Our main purpose is to present a unified approach to making inference for multi-level models from survey data, based on an extended weighted estimating equations approach and a weighted log composite likelihood approach (Sections 3 and 4). The proposed methods lead to asymptotically valid inferences on the variance component parameters even when the within-cluster sample sizes are small, unlike some of the existing methods summarized in Section 2. Limited simulation results are presented in Section 5.

2. TWO-LEVEL MODELS: PAST WORK

2.1. Two-level models

Multi-level (or hierarchical) models are extensively used in social sciences, education, health and other areas to analyze survey data with a hierarchical structure. Here we focus on two-level models associated with two-stage sampling of clusters (level 2): a sample, s, of level 2 units is selected according to a specified design and then a sample, $s(i)$, of elements (or level 1 units) is selected from each sampled level 2 unit i according to another specified design. It is assumed that the model matches the design hierarchy, but in practice design hierarchical structure may be different from that of the model hierarchical structure (Rao and Roberts, 1998).

Let N be the number of level 2 units in the population and M_i be the number of level 1 units in the level 2 unit i. A two-level super-population model is given by

$$y_{ij} \mid x_{ij}, v_i \sim_{ind} f\left(y_{ij} \mid x_{ij}, v_i, \theta_1\right), j = 1, \ldots, M_i \; ; \; v_i \sim_{ind} f\left(v_i \mid \theta_2\right), i = 1, \ldots, N \qquad (1)$$

where y_{ij} and x_{ij} are the response and covariate values associated with element j within cluster i, v_i denotes level 2 random effect, and θ_1 and θ_2 denote the parameters associated with the two levels of the assumed model. Here $f\left(y_{ij} \mid x_{ij}, v_i, \theta_1\right)$ and $f\left(v_i \mid \theta_2\right)$ are specified density functions of y_{ij} given x_{ij} and v_i and of v_i, respectively. The model formulation (1) covers both linear two-level models and generalized linear two-level models. Under informative sampling of clusters and of elements within sampled clusters, the population model (1) may not hold for the sample. In that case, standard methods for multi-level models that ignore the design and assume model (1) holds for the sample can lead to asymptotically biased estimators of model parameters θ_1 and θ_2 (Pfeffermann et al., 1998).

Mean model. A simple nested error mean model that is often used in simulation studies related to two-level models, is given by

$$y_{ij} = \mu + v_i + e_{ij}; e_{ij} \sim_{iid} N\left(0, \sigma_e^2\right), v_i \sim_{iid} N\left(0, \sigma_v^2\right). \qquad (2)$$

Model (2) may be written in the form (1) as

$$y_{ij} \mid v_i \sim_{ind} N\left(\mu + v_i, \sigma_e^2\right), v_i \sim_{iid} N\left(0, \sigma_v^2\right), \theta_1 = \left(\mu, \sigma_e^2\right), \theta_2 = \sigma_v^2 \; . \qquad (3)$$

2.2 Point estimation

A "census" log-likelihood under the assumed two-level model (1) is given by

$$\log L(\theta) = \sum_{i=1}^{N} \log L_i(\theta) \equiv \sum_{i=1}^{N} l_i(\theta), \tag{4}$$

where

$$L_i(\theta) = \int \exp\left\{\sum_{j=1}^{M_i} \log f\left(y_{ij} \mid x_{ij}, v_i, \theta_1\right)\right\} f\left(v_i \mid \theta_2\right) dv_i \;, \tag{5}$$

see Asparouhov (2006) and Rabe-Hesketh and Skrondal (2006).

Let the sample consist of n clusters with m_i elements from sample cluster i. Let π_i and $\pi_{j|i}$ respectively denote the level 2 and level 1 inclusion probabilities associated with cluster i and element j within cluster i. Then the level 2 and level 1 weights are given by $w_i = \pi_i^{-1}$ and $w_{j|i} = \pi_{j|i}^{-1}$ respectively. Asparouhov (2006) and Rabe-Hesketh and Skrondal (2006) proposed a weighted sample log pseudo-likelihood obtained by replacing $\sum_{j=1}^{M_i}(.)$ in (4) by $\sum_{j \in s_i} w_{j|i}(.)$ and $\sum_{i=1}^{N}(.)$ in (4) by $\sum_{i \in s} w_i(.)$, where s denotes the sample of clusters and $s(i)$ denotes the sample of elements within clusters $i \in s$. It is given by

$$\log L_w(\theta) = \sum_{i \in s} w_i l_{wi}(\theta) \equiv l_w(\theta) \tag{6}$$

where $l_{wi}(\theta) = \log L_{wi}(\theta)$ and

$$L_{wi}(\theta) = \int \exp\left\{\sum_{j \in s(i)} w_{j|i} \log f\left(y_{ij} \mid x_{ij}, v_i, \theta_1\right)\right\} f\left(v_i \mid \theta_2\right) dv_i. \tag{7}$$

Maximizing the log pseudo likelihood $l_w(\theta)$, given by (6), we get a pseudo maximum likelihood (PML) estimator $\tilde{\theta}_w$. Computational details are discussed in Asparouhov (2006) and Rabe-Hesketh and Skrondal (2006). In the special case of linear two-level models, Pfeffermann et al. (1998) used an iterative generalized least squares method proposed by Goldstein (1986). Note that we need both level 1 and level 2 weights to compute $\tilde{\theta}_w$, unlike in the case of marginal models that require only the unconditional element weights $w_{ij} = w_i w_{j|i}$.

Consistency with respect to both design and model of the PML estimators of variance components in the model requires that both the number of sample clusters, n, and the within cluster sample sizes, m_i, are large, even in the linear case. Also, the relative bias of the estimators will be considerable when the m_i are small. To remedy this problem, several weight-scaling methods have been proposed in the literature. In particular, level 1 weight $w_{j|i}$ in (7) is scaled by a factor k_{1i} before maximizing the pseudo log-likelihood (9). We consider only two weight-scaling methods here, denoted A and A1 (Asparouhov 2006). Method A uses

$$k_{1i} = m_i \Big/ \sum_{j \in s(i)} w_{j|i} \tag{8}$$

In method A1, k_{1i} is the same as in method A but level 2 weights w_i are also scaled by the factor $k_{2i} = 1/k_{1i}$ to offset level 1 weight scaling. Asparouhov (2006) mentioned the use of accelerated EM algorithm for calculating the PML estimator $\tilde{\theta}_w$ with M plus 3: www.Statmodel.com: Muthen and Muthen, 1998-2005.

2.3 Variance estimation

Turning to variance estimation, Asparouhov (2006) proposed a Taylor linearization "sandwich" variance estimator of $\tilde{\theta}_w$. It is given by

$$v_L\left(\tilde{\theta}_w\right) = \left(l''_w\right)^{-1}\left\{\sum_{i\in s}\left(k_{2i}w_i\right)^2 l'_{wi}\left(l'_{wi}\right)^T\right\}\left(l''_w\right)^{-1}, \tag{9}$$

where l'_w and l''_w respectively denote the first derivative vector and the second derivative matrix of $l_w(\theta)$ evaluated at $\theta = \tilde{\theta}_w$. If the level 2 sampling fraction is small, then $v_L(\tilde{\theta}_w)$ tracks the variance of $\tilde{\theta}_w$ well, but not the MSE of $\tilde{\theta}_w$ if the relative bias of $\tilde{\theta}_w$ is large.

Kovacevic et al. (2006) studied bootstrap variance estimators for $\tilde{\theta}_w$. They considered two options: options 1 and 2. In option 1, level 2 bootstrap weights $w_i(b)$, based on the Rao, Wu and Yue (1992) method, are used and level 1 weights are not changed, i.e., $w_{j|i}(b) = w_{j|i}$, where $b = 1,\ldots,B$ denote the B bootstrap samples. For option 2, the Rao, Wu and Yue (1992) bootstrap method is applied to both level 1 and level 2, and the level 1 bootstrap weights are rescaled. Replacing the weights w_i and $w_{j|i}$ by $w_i(b)$ and $w_{j|i}(b)$ in (6) and (7), bootstrap PML estimators $\tilde{\theta}_w(b), b = 1,\ldots,B$ are obtained and the resulting bootstrap variance estimator is given by

$$v_{Boot}(\tilde{\theta}_w) = \frac{1}{B}\sum_{b=1}^{B}\left\{\tilde{\theta}_w(b) - \tilde{\theta}_w\right\}\left\{\tilde{\theta}_w(b) - \tilde{\theta}_w\right\}^T. \tag{10}$$

A simulation study, based on the simple mean model (2) showed that option 1 may lead to underestimation of the within cluster variance component σ_e^2. Option 2 performed better than option 1. Grilli and Pratesi (2004) studied an alternative bootstrap method for variance estimation.

3. WEIGHTED ESTIMATING EQUATIONS: PROPOSED METHOD

3.1 Point estimation

We first illustrate the proposed weighted estimating equations approach, using the simple mean model (2). Here our interest is to estimate $\theta = \left(\mu, \sigma_v^2, \sigma_e^2\right)^T$ from a two-stage cluster sampling design with level 1 and level 2 weights given by $w_{j|i}$ and w_i respectively. We have chosen the following three estimating functions (EF) for this purpose:

$$u_1\left(y_{ij}, \theta\right) = y_{ij} - \mu,$$
$$u_2\left(y_{ij}, \theta\right) = \left(y_{ij} - \mu\right)^2 - \left(\sigma_v^2 + \sigma_e^2\right)$$
$$u_3\left(y_{ij}, y_{ik}\right) = \left(y_{ij} - y_{ik}\right)^2 - 2\sigma_e^2,\ j \neq k.$$

The resulting weighted estimating equations (WEE) are given by

$$\hat{U}_{1w}(\theta) = \sum_{i \in s} w_i \sum_{j \in s(i)} w_{j|i} u_1\left(y_{ij}, \theta\right) \equiv \sum_{i \in s} w_i \hat{U}_{1wi}(\theta) = 0 \tag{11}$$

$$\hat{U}_{2w}(\theta) = \sum_{i \in s} w_i \sum_{j \in s(i)} w_{j|i} u_2\left(y_{ij}, \theta\right) \equiv \sum_{i \in s} w_i \hat{U}_{2wi}(\theta) = 0 \tag{12}$$

$$\hat{U}_{3w}(\theta) = \sum_{i \in s} w_i \sum_{j \neq k \in s(i)} w_{jk|i} u_3(y_{ij}, y_{ik}, \theta) \equiv \sum_{i \in s} w_i \hat{U}_{3wi}(\theta) = 0, \tag{13}$$

where $w_{jk|i} = \pi_{jk|i}^{-1}$ and $\pi_{jk|i}$ denotes the joint inclusion probability for level 1 units j and k within level 2 unit i. The WEE estimator of θ is obtained by solving (14) – (16). For the mean model, we obtain explicit solutions to WEE given by

$$\hat{\mu}_w = \left(\sum_{i \in s} \sum_{j \in s(i)} w_{ij} y_{ij}\right) \Big/ \sum_{i \in s} \sum_{j \in s(i)} w_{ij} \equiv \bar{y}_w \tag{14}$$

$$\hat{\sigma}_{vw}^2 = \sum_{i \in s} \sum_{j \in s(i)} w_{ij} \left(y_{ij} - \bar{y}_w\right)^2 \Big/ \sum_{i \in s} \sum_{j \in s(i)} w_{ij} - \hat{\sigma}_{ew}^2 \tag{15}$$

$$\hat{\sigma}_{ew}^2 = \sum_{i \in s} w_i \sum_{j < k \in s(i)} w_{jk|i} \left(y_{ij} - y_{ik}\right)^2 \Big/ \left(2 \sum_{i \in s} w_i \sum_{j < k \in s(i)} w_{jk|i}\right), \tag{16}$$

where $w_{ij} = w_i w_{jk|i}$.

We note that $\hat{U}_{tw}(\theta), t = 1, 2, 3$ are estimating functions with zero expectation with respect to the design and the model, i.e. $E_m E_p \left\{\hat{U}_{tw}(\theta)\right\} = 0$. Using this result and some regularity conditions, it can be shown that the WEE estimator $\hat{\theta}_w = \left(\hat{\mu}_w, \hat{\sigma}_{vw}^2, \hat{\sigma}_{ew}^2\right)^T$ is design-model consistent for θ as the number of level 2 units in the sample, n, increases, even when the level 1 within cluster sample sizes, m_i, are small. This property does not necessarily hold for the estimators presented in Section 2. The proposed method, however, requires the within-cluster joint inclusion probabilities $\pi_{jk|i}$. The latter are readily available for simple random or stratified random sampling within clusters, or when the within cluster sampling fraction is small. Also several good approximations to $\pi_{jk|i}$ when sampling within clusters is based on unequal probability sampling are also available, and those approximations depend only on the marginal inclusion probabilities $\pi_{j|i}$ (Haziza, Mecatti and Rao, 2008).

The choice of estimating functions (11)-(13) is not necessarily unique. For example, we could replace the previous $u_2\left(y_{ij}, \theta\right)$ by $\tilde{u}_2\left(y_{ij}, y_{ik}, \theta\right) = \left(y_{ij} - \mu\right)\left(y_{ik} - \mu\right) - \sigma_v^2$ in (12) and retain (11) and (13). The resulting WEE estimator of θ is also design-model consistent as the number of level 2 units increases.

Korn and Graubard (2003) used an alternative approach for the mean model which has some similarities with the proposed approach. Under this approach, "census parameters", S_e^2 and S_v^2 are first obtained by assuming that the model holds for the finite population. Survey weighted estimators $\hat{S}_{ew}^2$ and $\hat{S}_{vw}^2$ of the census parameters are then obtained, assuming M_i is known for the sampled clusters. The estimator $\hat{S}_{ew}^2$ is given by

$$\hat{S}_{ew}^2 = \left[\frac{1}{2} \sum_{i \in s} \left(M_i - 1\right) w_i \left\{\sum_{j < k \in s(i)} w_{jk|i} \left(y_{ij} - y_{ik}\right)^2 \Big/ \sum_{j < k \in s(i)} w_{jk|i}\right\}\right] \left\{\sum_{i \in s} \left(M_i - 1\right) w_i\right\}^{-1}, \tag{17}$$

assuming m_i>1 for all sampled clusters. Note that (17) requires the joint inclusion probabilities $\pi_{jk|i}$ as in the proposed method, but it induces ratio bias when the within-cluster sample sizes are small unlike our method. Expression for $\hat{S}^2_{vw}$ is more complicated and we refer the reader to Korn and Graubard (2003) for the relevant formula.

3.2 Variance estimation

A Taylor linearization sandwich variance estimator of the WEE estimator $\hat{\theta}_w$ can be obtained along the lines of the variance estimator (9), provided the level 2 sampling fraction is small. Let $\hat{U}_w(\theta)$ be the column vector with components $\hat{U}_{1w}(\theta),\hat{U}_{2w}(\theta)$ and $\hat{U}_{3w}(\theta)$ and similarly $\hat{U}_{wi}(\theta)$ be the column vector with components $\hat{U}_{1wi}(\theta),\hat{U}_{2wi}(\theta),\hat{U}_{3wi}(\theta)$. Then the linearization variance estimator is given by

$$v_L\left(\hat{\theta}_w\right)=\left(\hat{U}'_w\right)^{-1}\left(\sum_{i\in s}w_i^2\hat{U}_{wi}\hat{U}_{wi}^T\right)\left(\hat{U}'_w\right)^{-1}, \qquad (18)$$

where $\hat{U}_w$ and $\hat{U}'_w$ denote $\hat{U}_{wi}(\theta)$ evaluated at $\theta=\hat{\theta}_w$ and the first derivative $\hat{U}'_{wi}(\theta)$ evaluated at $\theta=\hat{\theta}_w$, respectively. Properties of the variance estimator (18) will be studied in a more extensive paper under preparation. Bootstrap variance estimation along the lines of Kovacevic et al (2006) requires further study. But further complications might arise because of the presence of joint inclusion probabilities $\pi_{jk|i}$.

4. WEIGHTED LOG COMPOSITE LIKELIHOOD: A UNIFIED APPROACH

In this section we briefly describe a unified approach applicable to both linear and generalized linear multi-level models. This approach is based on the concept of composite likelihood which has become popular in the mainstream literature to handle clustered or spatial data (see e.g., Lindsay 1988, Lele and Taper 2002 and Cox and Reid 2004). A univariate marginal composite likelihood is obtained by multiplying the likelihood contributions from the individual observations within clusters. Similarly a pair-wise marginal composite likelihood is obtained by multiplying the likelihood contributions from all the distinct pairs within clusters. When the super-population model holds for the sample, then we can obtain parameter estimators by maximizing the two composite likelihoods separately. Here we extend this approach to handle informative designs by obtaining weighted estimating equations that require only the marginal weights $w_{j|i}$ and the pair wise weights $w_{jk|i}$, as in Section 3.

We obtain a weighted log univariate marginal composite likelihood as

$$l_{wC1}(\theta)=\sum_{i\in s}w_i\sum_{j\in s(i)}w_{j|i}\log f\left(y_{ij}\mid\theta\right) \qquad (19)$$

where $f\left(y_{ij}\mid\theta\right)$ is the marginal density function of y_{ij} and θ is the vector of model parameters. Similarly, a weighted log pair wise marginal composite likelihood is obtained as

$$l_{wC2}(\theta)=\sum_{i\in s}w_i\sum_{j<k\in s(i)}w_{jk|i}\log f\left(y_{ij},y_{ik}\mid\theta\right) \qquad (20)$$

where $f\left(y_{ij},y_{ik}\mid\theta\right)$ is the joint density of y_{ij} and y_{ik} .We then solve the weighted composite score equations

$$U_{wCt}(\theta)=\partial l_{wCt}(\theta)/\partial\theta=0, t=1,2 \qquad (21)$$

to get the weighted composite likelihood estimator of θ. If some of the elements of θ are common between the two weighted equations, then one can take the average of the two estimators (Fieuws and Verbeke, 2006) as the estimator for those elements. We plan to study the properties of the proposed weighted estimators in a future paper for both linear and generalized linear two level models and other extensions including missing data problems.

For the linear mixed models one could replace the bivariate density function by a univariate density function of the differences $z_{ijk} = y_{ij} - y_{ik}$ to simplify the method (Lele and Taper, 2002). For the simple mean model we have $z_{ijk} \sim N\left(0, 2\sigma_e^2\right)$. By reparametrizing θ as $\varphi = \left(\mu, \sigma^2, \sigma_e^2\right)$ where $\sigma^2 = \sigma_v^2 + \sigma_e^2$ and noting that $y_{ij} \sim N\left(\mu, \sigma^2\right)$, we see that the parameters of the two univariate densities are distinct and we solve the resulting weighted composite score equations separately to get an estimator of φ and in turn an estimator of θ. For the mean model, it can be easily shown that the above approach gives weighted estimating equations identical to (11)-(13) and hence identical estimators of the components of θ.

Turning to variance estimation, Taylor linearization variance estimators similar to (18) can be used provided that the level 2 sampling fraction is small. We plan to study variance estimation for the weighted composite likelihood estimators in a more detailed paper.

5. SIMULATION STUDY

We conducted a small simulation study on the performance of the proposed WEE estimators under the simple nested error mean model (2), using $\mu = 0.5, \sigma_v^2 = 0.5$ and $\sigma_e^2 = 2.0$. We confined our study to the special case of $N = n = 100$ clusters and equal size clusters, $M_i = M = 100$, with equal sample sizes $m_i = m = 5$. In this special case, clusters (level 2 units) are equivalent to strata because all clusters are sampled. We used the Rao-Sampford probability proportional to size (PPS) sampling without replacement with specified size measures z_{ij} to select $m_i = m = 5$ level 1 units from each level 2 unit (Rao 1965 and Sampford 1967). Following Asparouhov (2006), we considered both invariant and non-invariant selections as defined below. For invariant selection, we take

$$z_{ij} = \left(1 + \exp\left[-0.5\left\{e_{ij}/\alpha + e_{ij}^{*}\left(1-\alpha^{-2}\right)^{1/2}\right\}\right]\right)^{-1}, \qquad (22)$$

where e_{ij}^{*} is independent of e_{ij} but with the same distribution, $N\left(0, \sigma_e^2 = 2.0\right)$. In the case of non-invariant selection, we replaced e_{ij} and e_{ij}^{*} in (22) by $v_i + e_{ij}$ and $v_i^{*} + e_{ij}^{*}$ respectively, where v_i^{*} is independent of v_i but with the same distribution $N\left(0, \sigma_v^2 = 0.5\right)$. We considered four values of α in (22): $\alpha = 1,2,3,\infty$, where $\alpha = \infty$ corresponds to non-informative sampling within each stratum, $\alpha = 1$ corresponds to the most informative sampling and informativeness decreases as α increases. Asparouhov (2006) studied the case of most informative sampling only.

We used the design-model (pm) approach to simulate $R = 1{,}000$ samples for each specified α and separately for invariant and non-invariant selections. Under this approach, we generated a population with $N = 100$ and $M_i = M = 100$ from the model and then selected a sample of $m_i = 5$ units from each stratum (cluster) according to the size measures z_{ij}. The two-step process was repeated $R = 1{,}000$ times to simulate $1{,}000$ samples.

From each sample, we computed the estimates of μ, σ_v^2 and σ_e^2 using REML, weighted scaling methods A and A1, the proposed WEE method and the alternative method of Korn and Graubard (abbreviated KG). Biases and variances of the estimators were computed from the $1{,}000$ estimates. Performance of alternative estimators is judged using two performance measures: Bias ratio = BR = (Bias)/ (square root of variance) and relative root mean squared error = RRMSE = (square root of MSE)/ (true parameter value). Tables 1, 2 and 3 respectively report BR values of the estimators of μ, σ_v^2 and σ_e^2. RRMSE values the estimators of μ, σ_v^2 and σ_e^2 are reported in Tables 4, 5 and 6 respectively.

Bias ratio

Table 1 reports bias ratio (%) of the estimators of μ based on REML, A, A1, KG and WEE. Note that in the case of μ, estimators A1, KG and WEE are identical. Results in Table 1 show that BR is similar for invariant and non-invariant selections and that BR of REML and A decreases as α increases. Further, REML leads to large bias under informative sampling, even for $\alpha = 3$; for example, BR for REML ranges from 160% to 493% under invariant selection. Method A also leads to significant BR under informative sampling (40% to 117%). On the other hand, BR of WEE, A1 and KG does not depend on α and it is essentially negligible ($|BR| < 4\%$). Under non-informative sampling, REML performs well as expected ($|BR| < 3\%$).

Table 1 - Bias ratio (%) of estimators of μ

α	**Invariant**			**Non-invariant**		
	REML	**A**	**A1/WEE/KG**	**REML**	**A**	**A1/WEE/KG**
1	492.8	114.1	-0.5	533.6	116.7	-0.1
2	241.2	60.6	3.2	238.4	57.6	0.9
3	160.0	40.0	2.1	158.6	40.5	3.1
∞	-2.2	-2.1	-1.7	1.8	2.8	2.8

Turning to BR of the estimators of σ_v^2, Table 2 shows that BR of REML is not affected by α under invariant selection, but not under non-invariant selection. In the latter case, REML leads to serious underestimation for $\alpha = 1$ (BR= - 80%) but |BR| decreases as α increases. Table 2 also shows that methods A and A1 perform poorly in terms of BR even under non-informative sampling. KG did not perform well for $\alpha = 1$ (BR=45% under invariant selection and BR=26% under non-invariant selection). On the other hand, WEE performed well for all values of α (BR ranging from -10.6% to -6%) although underestimation is consistent across values of α.

Table 2 - Bias ratio (%) of estimators of σ_v^2

α	**REML**	**A**	**A1**	**WEE**	**KG**
Invariant Selection					
1	-3.1	86.8	86.8	-8.3	44.8
2	-1.4	38.8	41.6	-10.0	8.8
3	-3.3	29.4	31.6	-10.6	1.5
∞	-1.0	25.8	28.8	-8.7	-0.9
Non-invariant Selection					
1	-79.9	71.0	86.2	-6.0	26.3
2	-18.9	40.1	48.1	-6.5	6.8
3	-9.9	30.0	35.0	-8.3	0.6
∞	1.0	26.5	29.0	-6.4	0.4

Table 3 reports BR values of the estimators of σ_e^2 and it shows that BR values are similar for invariant and non-invariant selections, as in the case of μ. REML and KG lead to serious underestimation when $\alpha = 1$ (BR= -160% for REML and BR= - 110% for KG), but |BR| decreases as α increases and becomes negligible for $\alpha = \infty$. Estimators A and A1 perform poorly for all values of α including $\alpha = \infty$. On the other hand, WEE performs well for all value of α with |BR|<5%.

Table 3 - Bias ratio (%) of estimators of σ_e^2

α	REML	A	A1	WEE	KG
Invariant Selection					
1	-151.9	-171.2	-99.1	0.8	-106.0
2	-33.6	-60.9	-47.0	3.9	-23.6
3	-15.8	-44.1	-40.3	-0.0	-10.9
∞	-1.0	-31.5	-33.7	0.7	-0.7
Non-invariant Selection					
1	-156.1	-178.2	-103.7	-0.4	-108.4
2	-39.3	-68.3	-55.5	-2.9	-29.4
3	-18.1	-49.5	-46.9	-4.5	-14.4
∞	-2.1	-33.2	-36.2	-1.4	-1.2

Relative root mean squared error (%)

Table 4 shows that RRMSE values for estimators of μ are similar for invariant and non-invariant selections and that RRMSE of REML and A decreases as α increases. For informative sampling with $\alpha = 1$, RRMSE for REML is large relative to RRMSE for WEE (A1 and KG) due to large BR. For example, RRMSE=91% for REML compared to RRMSE=20% for WEE. As expected, REML has the smallest RRMSE under non-informative sampling, but the increase in RRMSE for the other methods is quite small. Also, RRMSE of WEE (A1 and KG) does not depend on α.

Table 4 - Relative root mean squared error (%) of estimators of μ

α	Invariant			Non-invariant		
	REML	A	A1/WEE/KG	REML	A	A1/WEE/KG
1	90.6	29.0	20.0	90.5	29.0	20.3
2	48.7	22.4	19.4	48.3	22.4	19.8
3	35.4	20.7	19.6	35.7	20.9	20.0
∞	18.9	19.1	19.4	19.1	19.5	19.7

Turning to RRMSE of estimators of σ_v^2, Table 5 shows that REML performs well for all α under invariant selection due to small BR in this case. We also note that KG and WEE are comparable in terms of RRMSE for all values of α. Table 5 also shows that A and A1 lead to somewhat larger RRMSE for $\alpha = 1$ (42% for A1 and 38% for A compared to 31% for WEE).

Table 5 - Relative root mean squared error (%) of estimators of σ_v^2

α	REML	A	A1	WEE	KG
Invariant Selection					
1	24.3	37.5	41.7	29.2	30.7
2	26.5	29.8	31.3	29.8	28.3
3	26.2	28.4	29.4	28.9	27.6
∞	25.9	27.2	27.8	27.7	26.8
Non-invariant Selection					
1	29.4	34.3	42.4	30.6	28.9
2	24.5	27.5	30.1	27.2	25.8
3	25.7	27.7	28.9	28.1	26.8
∞	26.3	28.5	29.3	29.1	27.9

Table 6 gives RRMSE values of the estimators of σ_e^2 and we note that the values are similar for invariant and non-invariant selections. It also shows that RRMSE values are comparable for methods WEE, A, A1 and KG even though in terms of bias ratio A, A1 and KG performed poorly relative to WEE. This is due to larger variance for WEE compared to other methods. For example, in the case of invariant selection and $\alpha = 1$ we have the following variances for WEE, KG and REML: 0.0378, 0.0214 and 0.0170 with corresponding bias ratios (%) from Table 3: -0.8, -151.9, and -106.0.

Table 6 - Relative root mean squared error (%) of estimators of σ_e^2

α	REML	A	A1	WEE	KG
Invariant Selection					
1	11.9	13.1	10.5	9.7	10.7
2	7.2	8.2	8.1	8.3	7.5
3	7.2	7.9	8.0	8.1	7.5
∞	7.2	7.6	7.7	7.8	7.4
Non-invariant Selection					
1	11.7	13.0	10.3	9.3	10.5
2	7.3	8.3	8.1	8.0	7.5
3	7.0	7.8	7.8	7.7	7.3
∞	7.0	7.5	7.7	7.8	7.3

6. CONCLUDING REMARKS

In this paper, we have proposed unified design-based approaches to making inferences for multi-level models from complex survey data. The proposed methods are asymptotically valid even when the sample sizes within sampled clusters (level 2 units) are small, unlike some of the existing methods, but knowledge of joint inclusion probabilities within sampled clusters is required. Often it may be possible to treat the sample within clusters as drawn with replacement because of small sampling fractions within clusters. Also, excellent approximations to joint inclusion probabilities, depending only on the marginal inclusion probabilities, are also available when the sampling fractions are not small (Haziza et al., 2008). We plan to study the accuracy of such approximations in a future study.

In the simulation study, we assumed that all the level 2 units are sampled, for simplicity. We plan to extend the simulation study to cover the case of sampling the level 2 units as well as the level 1 units within the sampled level 2 units. We also propose to study properties of variance estimators, including the Taylor linearization variance estimator (18) and bootstrap variance estimators.

We also plan to study the unified weighted log-composite likelihood approach in detail, including its theoretical properties, and apply it to handle generalized linear multi-level models, including logistic linear multi-level models for binary responses.

REFERENCES

Asparouhov, T. (2006). Generalized multi-level modeling with sampling weights. *Communications in Statistics – Theory and Methods*, **35**, 439-460.

Beaumont, J-F. and Charest, A-S. (2010). Bootstrap variance estimation with survey data when estimating model parameters. Unpublished report (courtesy of the authors).

Cox, D.R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, **91**, 729-737.

Fieuws, S. and Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, **62**, 424-431.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, **73**, 43-56.

Grille, L. and Pratesi, M. (2004). Weighted estimation in multilevel ordinal and binary models in the presence of informative sampling designs. *Survey Methodology*, **30**, 93-103.

Haziza, D., Mecatti, F. and Rao, J.N.K. (2008). Evaluation of some approximate variance estimators under the Rao-Sampford unequal probability sampling design. *Metron*, **66**, 91-108.

Korn, E.L. and Graubard, B.I. (2003). Estimating variance components using survey data. *Journal of the Royal Statistical Society B*, **65**, 175-190.

Kovacevic, M.S., Rong, H. and You, Y. (2006). Bootstrapping for variance estimation in multi-level models fitted to survey data. *Proceedings of ASA Section on Survey Research Methods*, pp. 3260-3269, American Statistical Association.

Lele, S. and Taper, M.L. (2002). A composite likelihood approach to (co)variance components estimation. *Journal of Statistical Planning and Inference*, **103,** 117-125.

Lindsay, B.G. (1988). Composite likelihood methods. In *Statistical Inference from Stochastic Processes*, Ed. N. U. Prabhu, pp. 221-239, Providence: American Mathematical Society.

Pfeffermann, D. and Sverchkov, M. (2003). Fitting generalized linear models under informative sampling. In *Analysis of Survey Data* (Eds. R. Chambers and C.J. Skinner) 175-196, Wiley, Chichester.

Pfeffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H. and Rasbash, J. (1998). Weighting for unequal selection probabilities in multi-level models. *Journal of the Royal Statistical Society B*, **60**, 23-56.

Rao, J.N.K. (1965). On two simple schemes of unequal probability sampling without replacement. *Journal of the Indian Statistical Association*, **3**, 173-180.

Rao, J.N.K. and Roberts, G. (1998). Discussion on the papers by Firth and Bennett and Pfeffermann et al. *Journal of the Royal Statistical Society B*, **60**, 50-51.

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, **18**, 209-217.

Rao, J.N.K., Hidiroglou, M., Yung, W. and Kovacevic, M. (2010). Role of weights in descriptive and analytical inferences from survey data: an overview. *Journal of the Indian Society of Agricultural Statistics*, **64**, 129-135.

Rabe-Hesketh, S. and Skrondal, A. (2006). Multilevel modeling of complex survey data. *Journal of the Royal Statistical Society A*, **169**, 805-827.

Sampford, M.R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika*, **54,** 499-513.