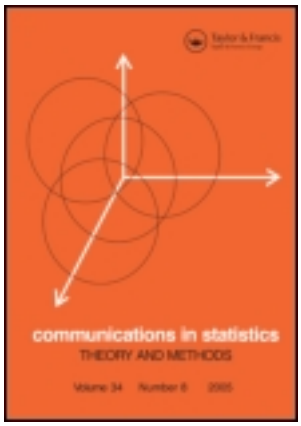

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

Fitting Variance Components Model and Fixed Effects Model for One-Way Analysis of Variance to Complex Survey Data

Abdulhakeem A. H. Eideh [a]
[a] Department of Mathematics, Al-Quds University, Palestine, Jerusalem

Version of record first published: 25 Jul 2012

To cite this article: Abdulhakeem A. H. Eideh (2012): Fitting Variance Components Model and Fixed Effects Model for One-Way Analysis of Variance to Complex Survey Data, Communications in Statistics - Theory and Methods, 41:16-17, 3278-3300

To link to this article: http://dx.doi.org/10.1080/03610926.2012.692425

Communications in Statistics—Theory and Methods, 41: 3278–3300, 2012

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610926.2012.692425

Taylor & Francis
Taylor & Francis Group

Fitting Variance Components Model and Fixed Effects Model for One-Way Analysis of Variance to Complex Survey Data

ABDULHAKEEM A. H. EIDEH

Department of Mathematics, Al-Quds University, Palestine, Jerusalem

Under complex survey sampling, in particular when selection probabilities depend on the response variable (informative sampling), the sample and population distributions are different, possibly resulting in selection bias. This article is concerned with this problem by fitting two statistical models, namely: the variance components model (a two-stage model) and the fixed effects model (a single-stage model) for one-way analysis of variance, under complex survey design, for example, two-stage sampling, stratification, and unequal probability of selection, etc. Classical theory underlying the use of the two-stage model involves simple random sampling for each of the two stages. In such cases the model in the sample, after sample selection, is the same as model for the population; before sample selection. When the selection probabilities are related to the values of the response variable, standard estimates of the population model parameters may be severely biased, leading possibly to false inference. The idea behind the approach is to extract the model holding for the sample data as a function of the model in the population and of the first order inclusion probabilities. And then fit the sample model, using analysis of variance, maximum likelihood, and pseudo maximum likelihood methods of estimation. The main feature of the proposed techniques is related to their behavior in terms of the informativeness parameter. We also show that the use of the population model that ignores the informative sampling design, yields biased model fitting.

Keywords Fixed effects model; Informative sampling; Maximum likelihood estimation; Pseudo maximum likelihood; Sample distribution; Variance components model.

1. Introduction

In classifying data in terms of factors and their levels, the feature of interest is the extent to which different levels of a factor affect the variable of interest. The effects

Received January 5, 2011; Accepted May 7, 2012
Address correspondence to Abdulhakeem A. H. Eideh, Department of Mathematics, Chairman, Faculty of Science and Technology, Al-Quds University, Abu-Dies campus, Palestine P.O. Box 20002, Jerusalem; E-mail: msabdul@science.alquds.edu

of a factor are always one or other of the two kinds. The first kind is fixed effects, which is the effects attributable to a finite set of levels of a factor that occur in the data and exist because the interest in them. Models in which the only effects are fixed effects are called fixed effects models. The second kind of effects is random effects. These are attributable to (usually) infinite set of levels of a factor in which only a random sample are deemed to occur in the data. Models in which the only effects are random are called random effects models (or variance components models). For future discussion on the analysis of these models in nonsurvey context, we refer the reader to Searle et al. (1992).

In survey context, multi-stage, two-stage, and single-stage population models and the corresponding sampling methods are frequently used in the health and social sciences for the modeling of hierarchically structured populations. Classical theory underlying the use of two-stage sampling method involves simple random sampling for each of the two stages or fixed unequal probabilities of selection at one or more of the two stages. In such cases, the model in the sample is the same as the mode for the population. When the selection probabilities are related to the values of the response variable, the sample design is defined as informative. This may lead to the model holding for the sample being different from the model holding in the population, resulting in selection bias. Thus, standard estimates of the population model parameters may be severely biased, leading possibly to false inference, see for example, Pfeffermann et al. (1998a). Consider an education study of pupils' proficiencies with schools as the first stage sampling units and pupils as second stage sampling units. And in addition, suppose that the schools are selected with probabilities proportional to their sizes (number of pupils). If the size of the school is related to the school average of pupils' proficiencies, say the large schools are mostly in areas with low proficiencies, and the size of the school is not included among the model covariates, therefore sample of schools will tend to contain large schools with low proficiencies, and hence no longer represent the population of schools.

Pfeffermann et al. (2006) pointed out that "a possible way to deal with the problem of informative sampling is to include among the model covariates the design variables that define the selection probabilities at the various levels." However, this paradigm is often not practical. Firstly, not all design variables used for the sample selection may be known or accessible to the analysts, or that there may be too many of them, making the fitting and validation of such models formidable. Secondly, by including the design variables among the model covariates, the resulting model may be no longer of scientific interest. This is not necessarily a problem if the model is fitted for prediction purposes but is clearly not acceptable when the purpose of the analysis is to study the structural relationships between the outcome variable and covariates of interest." Furthermore, Korn and Graubard (1999, pp. 179–180) pointed out that "the approach to modeling the sampling design in the general case is as follows: Start with a model that would be used if a simple random sample of the population were being analyzed. If the inefficiency of weighted estimation for the primary parameter of interest is unacceptably large, then consider augmenting the model with variables used in the construction of the sample weights. Such survey-design variables are those defining the sampling strata, the nonresponse weighting cells, and the poststratification adjustment cells. These survey-design variables are included in the model provided that they do not lessen the interpretability of the primary parameter."

Pfeffermann et al. (1998b) considered two approaches to weighting iterative generalized least squares for multilevel models. The first approach uses the reciprocals of selection probabilities and follows the broad principles of the pseudo likelihood approach. The second approach scales the weights, that is it replaces each weight by, for example, the relative weight (that is divided by the sample mean of the weights), in order to improve the properties of the estimators and to simplify computation. Korn and Graubard (2003) proposed new estimators for variance components, some of which are approximately unbiased regardless of the sampling design. These estimators require knowledge of the joint inclusion probabilities of the observations. The small sample properties of the estimators are studied via simulation for the simple one-way random-effects model. Pfeffermann et al. (2006) considered a model-dependent approach for multilevel modeling that accounts for informative probability sampling of first- and lower-level population units. Their approach proposed consists of first extracting the hierarchical model holding for the sample data, given the selected sample, as a function of the corresponding population model and the first- and lower-level sample selection probabilities. Then followed by fitting the resulting sample model, using Bayesian methods. An important implication of the use of the model holding for the sample is that the sample selection probabilities feature in the analysis as additional data that possibly strengthen the estimators. A simulation experiment, carried out in order to study the performance of this approach and compare it to the use of "design-based" methods indicates that both approaches perform in general equally well in terms of point estimation. However, the model-dependent approach yields confidence/credibility intervals with better coverage properties. Another simulation study assesses the impact of misspecification of the models assumed for the sample selection probabilities. The use of maximum likelihood estimation is also considered. Jia et al. (2011) considered the performance of random effects model estimators under complex sampling designs. In particular, they derive analytical formulae for the bias in random effects ANOVA. Their approach include verifying the formulae by means of Monte Carlo simulations and use the expressions to examine the impact of sample size, the size of the intraclass correlation coefficient (ICC), and the sampling design on the estimators' performance. They also consider the controversial issue of scaling and the extension to second-order weights.

None of the above studies consider the fitting of a variance components model or a fixed effects model for one-way analysis of variance to complex survey data under an informative probability sampling design. This article is devoted to the fitting of two statistical models for complex survey data under informative probability sampling design, namely; the two-stage and single-stage population models for one-way classification. Depending on the fact that, in this work, the design variables used for the sample selection are not included in the models.

The research material in this article is structured as follows. In Sec. 2, we outline the main features of sample likelihood under informative sampling. Section 3 considers the variance component models under informative sampling. Section 4 contains fixed effects models for one-way classification under informative sampling. Section 5 is devoted to the estimation of variance components and fixed effects models parameters. In Sec. 6, we present a comparison between different estimators. Finally, Sec. 7 provides a discussion of the results.

2. Sample Likelihood under Informative Sampling

Let $U = \{1, \ldots, N\}$ denote a finite population consisting of N units. Let y be the study variable of interest and let y_i be the value of y for the ith population unit. We consider the population values $y_1, \ldots, y_N$ as random variables, that are independent realizations from a distribution with a probability density function (pdf) $f_p(y_i \mid \boldsymbol{\theta})$, indexed by a vector of parameters, $\boldsymbol{\theta}$. Let $\mathbf{z} = (z_1, \ldots, z_N)'$ be the values of known design variables, used for the sample selection process. In what follows, we consider a sampling design with selection probabilities $\pi_i = \Pr(i \in s)$, and sampling weight $w_i = 1/\pi_i$; $i = 1, \ldots, N$. In practice, the π_i's may depend on the population values $(\mathbf{y}, \mathbf{z})$, where $\mathbf{y} = (y_1, \ldots, y_N)'$. We express this dependence by: $\pi_i = \Pr(i \in s \mid \mathbf{y}, \mathbf{z})$ for all units $i \in U$. The sample s consists of the subset of U selected at random by the sampling scheme with inclusion probabilities $\pi_1, \ldots, \pi_N$. We assume probability sampling, such that $\pi_i = \Pr(i \in s)$ are strictly positive for all units $i \in U$. The sample distribution refers to the superpopulation distribution of the sample measurements as induced by the population model and the sample selection scheme with the selected sample of units held fixed.

Before defining the sample distribution mathematically, let us introduce the following notations: f_p and $E_p(\cdot)$ denote the pdf and the mathematical expectation of the population distribution, respectively, while f_s and $E_s(\cdot)$ denote the pdf and the mathematical expectation of the sample distribution, respectively. According to Pfeffermann et al. (1998a), the sample pdf of y_i is defined as:

$$\begin{aligned} f_s(y_i \mid \boldsymbol{\theta}, \gamma) &= f_p(y_i \mid \boldsymbol{\theta}, i \in \mathrm{s}) \\ &= \frac{\Pr(i \in s \mid y_i, \gamma) f_p(y_i \mid \boldsymbol{\theta})}{\Pr(i \in s \mid \boldsymbol{\theta}, \gamma)} \\ &= \frac{E_p(\pi_i \mid y_i, \gamma)\, f_p(y_i \mid \boldsymbol{\theta})}{E_p(\pi_i \mid \boldsymbol{\theta}, \gamma)}, \end{aligned} \tag{1}$$

where

$$E_p(\pi_i \mid \boldsymbol{\theta}, \gamma) = \int E_p(\pi_i \mid y_i, \gamma)\, f_p(y_i \mid \boldsymbol{\theta})\, dy_i$$

and γ is a parameter, from now on called "informativeness parameter," associated to sample measurement y_i in the model used to describe the selection procedure, i.e., in the conditional expectation of the sample inclusion probabilities, $E_p(\pi_i \mid y_i, \gamma) = \Pr(i \in s \mid y_i, \gamma)$. This previously defined parameter is called informativeness parameter, and if it is equal to zero, it follows that the sample design is free of y_i and then it is not informative.

Note that the sample pdf is different from the superpopulation pdf generating the finite population values, unless $\Pr(i \in s \mid y_i, \gamma) = \Pr(i \in s \mid \boldsymbol{\theta}, \gamma)$ for all possible values of y_i, in which case the sampling process is noninformative and can be ignored for purposes of inference. Also note that the sample distribution is a function of the population distribution and of the first order sample inclusion probabilities.

In practice, the conditional expectations of the sample inclusion probabilities $E_p(\pi_i \mid y_i, \gamma)$ are not known. Assuming that the data available to the analyst is $\{y_i, w_i; i \in s\}$, which is the case for secondary analysis. The question now that arises

is: how can we identify and estimate $E_p(\pi_i \mid y_i, \gamma)$ based only on the sample data? The following relationships, due to Pfeffermann and Sverchkov (1999), answer this question:

$$E_s(w_i \mid y_i, \gamma) = \left\{E_p(\pi_i \mid y_i, \gamma)\right\}^{-1} \tag{2}$$

$$E_s(w_i) = \left\{E_p(\pi_i)\right\}^{-1}. \tag{3}$$

Having derived the sample distribution, Pfeffermann et al. (1998a) proved that if the population measurements y_i are independent, then as $N \to \infty$ (with n fixed) the sample measurements are asymptotically independent. As a result, we can apply standard inference procedures to complex survey data by using the marginal sample distribution for each unit. But as mentioned in Pfeffermann et al. (1998a), when the conditional expectation of the first-order sample selection probabilities is exponential, for example, $E_p(\pi_i \mid y_i) = \exp(\gamma_0 + \gamma_1 y_i)$, the problem of identifiability arises. For more explanation, see (Eideh, 2010, Secs. 5 and 8). Also, there are cases where the number of parameters indexing the sample distribution (the parameters that index the conditional expectation of the first order sample selection probabilities and the parameters that characterize the population) is large. In these cases, based on the sample data $\{y_i, w_i; i \in s\}$, Pfeffermann et al. (1998a) proposed a two-step estimation method, which can be presented as follows:

Step 1. Estimate the parameter γ using Eq. (2). Denoting the resulting estimate of γ by $\hat{\gamma}$.

Step 2. Substitute $\hat{\gamma}$ in the sample log-likelihood function, with holding fixed the estimate of γ, and then maximize the resulting sample log-likelihood function with respect to the population parameters, $\boldsymbol{\theta}$:

$$l_{rs}(\boldsymbol{\theta}, \hat{\gamma}) = l_{srs}(\boldsymbol{\theta}) + \sum_{i=1}^{n} \ln E_p(\pi_i \mid y_i, \hat{\gamma}) - \sum_{i=1}^{n} \ln E_p(\pi_i \mid \boldsymbol{\theta}, \hat{\gamma}).$$

However, $\sum_{i=1}^{n} \ln E_p(\pi_i \mid y_i, \hat{\gamma})$ does not contain the population parameter, $\boldsymbol{\theta}$. Then using Eq. (3), we have:

$$l_{rs}(\boldsymbol{\theta}, \hat{\gamma}) = l_{srs}(\boldsymbol{\theta}) + \sum_{i=1}^{n} \ln E_s(w_i \mid \boldsymbol{\theta}, \hat{\gamma}),$$

where $l_{rs}(\boldsymbol{\theta}, \hat{\gamma})$ is the sample log-likelihood after substituting $\hat{\gamma}$ in the sample log-likelihood function and $l_{srs}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log\left\{f_p(y_i \mid \boldsymbol{\theta})\right\}$ is the classical log-likelihood.

For more discussion about the analysis of complex survey data under informative probability sampling design, see Pfeffermann and Sverchkov (1999, 2003), Eideh (2009), Eideh and Nathan (2009), Chambers and Skinner (2003), and Skinner (1994).

3. Variance Component Models under Informative Sampling

3.1. *Two-Stage Population Model*

Let $U = \{1, \ldots, N\}$ be a finite population of N primary sampling units (psu's), and $M_i; i = 1, \ldots, N$ be the number of secondary sampling units (ssu's) in the ith psu.

Let $y_{ij}; i = 1, \ldots, N; j = 1, \ldots, M_i$ be the value of the response variable y for the ssu j belonging to the psu i. The two-stage population model that includes a random intercept effect is given by:

$$\begin{aligned} &\text{First stage: } \mu_i = \mu + \eta_i; \quad i = 1, \ldots, N, \\ &\text{Second stage: } y_{ij} \mid \mu_i = \mu_i + e_{ij}; \quad j = 1, \ldots, M_i \end{aligned} \tag{4}$$

where e_{ij} and η_i are independent, and the population distributions of e_{ij} and η_i are $N(0, \sigma_e^2)$ and $N(0, \sigma_\mu^2)$, respectively.

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

For this model we have:

$$\begin{aligned} E_p\left(y_{ij}\right) &= \mu, \\ Var_p\left(y_{ij}\right) &= \sigma_\mu^2 + \sigma_e^2 \\ Cov_p\left(y_{ij}, y_{ik}\right) &= \sigma_\mu^2, j \neq k, \\ Cov_p\left(y_{ij}, y_{rk}\right) &= 0, i \neq r. \end{aligned} \tag{5}$$

The purpose of this article is to estimate the population mean μ and the variance components, σ_μ^2 and σ_e^2 when the sampling design for both of the two stages is informative.

Under the assumptions of the model given in Eq. (4), we can show that the population distribution of $\mathbf{y}_i$ is $N\left(\mu \mathbf{1}_{M_i}, \mathbf{V}_i\right)$ where $\mathbf{1}_{M_i} = (1, \ldots, 1)'$ is a vector of length M_i, $\mathbf{V}_i = \sigma_\mu^2 \mathbf{J}_{M_i} + \sigma_e^2 \mathbf{I}_{M_i}$, $\mathbf{J}_{M_i}$ is a square matrix of order M_i with every element equal one, and $\mathbf{I}_{M_i}$ is the identity matrix of order M_i. According to Searle et al. (1992, p. 79), the population pdf of $\mathbf{y}_i = \left(y_{i1}, \ldots, y_{iM_i}\right)'$ can be written as:

$$\begin{aligned} f_p\left(\mathbf{y}_i\right) &= \int f_p\left(y_{i1}, \ldots, y_{iM_i}, \mu_i\right) d\mu_i \\ &= \int f_p\left(\mu_i\right) \prod_{j=1}^{M_i} f_p\left(y_{ij} \mid \mu_i\right) d\mu_i \\ &= (2\pi)^{-0.5M_i} \left(\sigma_e^2\right)^{-0.5(M_i-1)} \left(M_i\sigma_\mu^2 + \sigma_e^2\right)^{-0.5} * \exp\left[-\frac{1}{2\sigma_e^2} \sum_{j=1}^{M_i} \left(y_{ij} - \mu\right)^2\right] \\ &\quad * \exp\left[\frac{\sigma_\mu^2}{\left(2\sigma_e^2\right)\left(M_i\sigma_\mu^2 + \sigma_e^2\right)} \left[\sum_{j=1}^{M_i} \left(y_{ij} - \mu\right)\right]^2\right]. \end{aligned} \tag{6}$$

This form plays an important role in estimation.

3.2. *Sample Design*

We assume a two-stage cluster sample 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 x_i and $z_{ij}; i = 1, \ldots, N; j = 1, \ldots, M_i$ 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_I of size n psu's (clusters) is selected with inclusion probabilities: $\pi_i = \Pr(i \in s_I \mid \mu_i, x_i)$ for all psu's $i \in U$. At the second stage a sample, s_i, of size m_i ssu's is selected from the ith selected psu with conditional inclusion probabilities: $\pi_{j|i} = \Pr\left(j \in s_i \mid i \in s, y_{ij}, z_{ij}\right)$, for all ssu j belonging to psu $i \in s_I$. In the following, we use only the conditional expectations of the inclusion probabilities $E_p(\pi_i \mid \mu_i)$ and $E_p\left(\pi_{j|i} \mid y_{ij}\right)$. Conditions for identification related to the presence of covariates at both the first and second level were discussed in Pfeffermann et al. (2006) and in Eideh and Nathan (2009).

3.3. *Sample Marginal Distributions of Response Measurements*

In order to obtain the sample marginal distribution of response measurements, and consequently the sample likelihood function, we need the sample distribution of the random effects. In addition, we need the sample distribution of the response variable given the random effect.

According to Eq. (1), the first stage sample distribution of the random effects μ_i is:

$$f_s(\mu_i) = \frac{E_p(\pi_i \mid \mu_i)}{E_p(\pi_i)} f_p(\mu_i).$$

Note that, for given $f_p(\mu_i)$, $f_s(\mu_i)$ is completely determined by specifying $E_p(\pi_i \mid \mu_i)$.

Pfeffermann et al. (1998a) and Skinner (1994) considered two possible models for the expectations of the first-stage sample inclusion probabilities, namely the polynomial and exponential models of the response variable. Eideh (2003) and Nathan and Eideh (2004) also considered the logit and probit models. As an illustration, in this article we shall consider only the exponential model:

$$E_p(\pi_i \mid \mu_i) = \exp(b_0 + b_1\mu_i), \tag{7}$$

where b_0 and b_1 are unknown parameters, to be estimated from the sample data, see Sec. 5.1.1., however it is easy to extend the results to the case where the conditional expectations of the inclusion probabilities are not of exponential form, for example linear, logit or probit. As pointed out by Skinner (1994), "this exponential approximation model for first order inclusion probabilities is appealing in the common situation where the sample selection is carried out in several stages so that the ultimate inclusion probabilities are the products of the selection probabilities at the various stages."

Now, under Eq. (7) and since the population distribution of μ_i is $N\left(\mu, \sigma_\mu^2\right)$, we can obtain:

$$\begin{aligned} f_s(\mu_i) = \frac{E_p(\pi_i \mid \mu_i)}{E_p(\pi_i)} f_p(\mu_i) &= \frac{\exp(b_0 + b_1\mu_i)}{\sqrt{2\pi\sigma_\mu^2} M_p(b_1)} \exp\left(-\frac{1}{2\sigma_\mu^2}(\mu_i - \mu)^2\right) \\ &= \frac{\exp\left(-\frac{1}{2\sigma_\mu^2}(\mu_i - \mu)^2 + b_1\mu_i - b_1\mu - \sigma_\mu^2 b_1^2/2\right)}{\sqrt{2\pi\sigma_\mu^2}} \\ &= \frac{\exp\left(-\frac{1}{2\sigma_\mu^2}\left[\left(\mu_i^2 - 2\mu_i\mu + \mu^2\right) - 2\sigma_\mu^2\left(b_1\mu_i - b_1\mu - \sigma_\mu^2 b_1^2/2\right)\right]\right)}{\sqrt{2\pi\sigma_\mu^2}} \end{aligned}$$

Completing the square in μ_i, we get:

$$f_s(\mu_i) = \left(2\pi\sigma_\mu^2\right)^{-0.5} \exp\left(-\frac{1}{2\sigma_\mu^2}\left(\mu_i - \left(\mu + b_1\sigma_\mu^2\right)\right)^2\right), \tag{8}$$

that is, the sample distribution of μ_i is $N\left(\mu + b_1\sigma_\mu^2, \sigma_\mu^2\right)$. Hence, the sample and population models belong to the normal distribution, but the mean of the sample model shifts by the constant $b_1\sigma_\mu^2$. For more information on sample distribution, see Pfeffermann et al. (1998a).

Note that if the informativeness parameter $b_1 = 0$, that is, the sampling design is noninformative, then the sample and population distributions of μ_i coincide, and in such cases, the sampling design is ignorable for statistical inference.

Similar to the sample distribution of random effects, the conditional sample pdf of y_{ij} given μ_i is given by:

$$f_s\left(y_{ij} \mid \mu_i\right) = \frac{E_p\left(\pi_{j\mid i} \mid \mu_i, y_{ij}\right)}{E_p\left(\pi_{j\mid i} \mid \mu_i\right)} f_p\left(y_{ij} \mid \mu_i\right).$$

If the population distribution of $y_{ij} \mid \mu_i$ is $N\left(\mu_i, \sigma_e^2\right)$ and

$$E_p\left(\pi_{j\mid i} \mid y_{ij}, \mu_i\right) = \exp\left(d_0 + d_1 y_{ij}\right), \tag{9}$$

where d_0 and d_1 are unknown parameter, to be estimated from the sample data, see Sec. 5.1.1, then similar to the procedure used in obtaining Eq. (8), we have:

$$f_s\left(y_{ij} \mid \mu_i\right) = \left(2\pi\sigma_e^2\right)^{-0.5} \exp\left(-\frac{1}{2\sigma_e^2}\left(y_{ij} - \left(\mu_i + d_1\sigma_e^2\right)\right)^2\right), \tag{10}$$

that is, the sample distribution of $y_{ij} \mid \mu_i$ is $N\left(\mu_i + d_1\sigma_e^2, \sigma_e^2\right)$.

Note that if the informativeness parameter $d_1 = 0$, that is, the sampling design is noninformative, then the sample and population distributions of $y_{ij} \mid \mu$ coincide, and in such cases, the sampling design is ignorable for statistical inferences.

Thus, based on Eqs. (8) and (10), the two-stage sample model is given by:

$$\begin{aligned} &\text{First stage: } \mu_i = \mu + b_1\sigma_\mu^2 + \eta_i; \quad i = 1, \ldots, n, \\ &\text{Second stage: } y_{ij} \mid \mu_i = \mu_i + d_1\sigma_e^2 + e_{ij}; \quad j = 1, \ldots, m_i, \end{aligned} \tag{11}$$

where e_{ij} and η_i are independent, and the sample distributions of e_{ij} and η_i are $N\left(0, \sigma_e^2\right)$ and $N\left(0, \sigma_\mu^2\right)$, respectively.

Now, we are interested in deriving the sample marginal distributions of the ith sampled cluster, whose sample measurement is $\mathbf{y}_i = \left(y_{i1}, \ldots, y_{im_i}\right)$, when the sample designs for both the first stage and second stage are informative. The sample model and moments of y_{ij}, under the two-stage sample design, depend on the sample model of the cluster-specific effects μ_i (first stage) and the sample model of $y_{ij} \mid \mu_i$ (second stage). Under Eqs. (7) and (9), and since $\left(y_{i1}, \ldots, y_{im_i}\right) \mid \mu_i; i = 1, \ldots, n$ are independent, therefore:

$$\begin{aligned} f_s\left(y_{i1}, \ldots, y_{im_i}\right) &= \int f_s\left(y_{i1}, \ldots, y_{im_i}, \mu_i\right) d\mu_i \\ &= \int f_s\left(\mu_i\right) \prod_{j=1}^{m_i} f_s\left(y_{ij} \mid \mu_i\right) d\mu_i \end{aligned}$$

Then, using Eqs. (6), (8), and (10), we can show that:

$$f_s\left(y_{i1},\ldots,y_{im_i}\right) = (2\pi)^{-0.5m_i}\left(\sigma_e^2\right)^{-0.5(m_i-1)}\left(m_i\sigma_\mu^2+\sigma_e^2\right)^{-0.5} * \exp\left[-\frac{1}{2\sigma_e^2}\left[\sum_{j=1}^{m_i}\left\{y_{ij}-\left(\mu+b_1\sigma_\mu^2+d_1\sigma_e^2\right)\right\}^2\right]\right] * \exp\left[\frac{\sigma_\mu^2}{\left(2\sigma_e^2\right)\left(m_i\sigma_\mu^2+\sigma_e^2\right)}\left[\sum_{j=1}^{m_i}\left\{y_{ij}-\left(\mu+b_1\sigma_\mu^2+d_1\sigma_e^2\right)\right\}\right]^2\right]. \quad (12)$$

This equation can be written in matrix form as:

$$f_s\left(y_{i1},\ldots,y_{im_i}\right) = \frac{\exp\left(-0.5\left(\mathbf{y}_i-\mathbf{1}_{m_I}\mu^*\right)'\mathbf{V}_i^{-1}\left(\mathbf{y}_i-\mathbf{1}_{m_I}\mu^*\right)\right)}{(2\pi)^{0.5m_i}\left|\mathbf{V}_i\right|^{0.5}},$$

where, the sample distribution of $\mathbf{y}_i=\left(y_{i1},\ldots,y_{im_i}\right)'$ is $N\left(\mathbf{1}_{m_i}\mu^*,\mathbf{V}_i\right)$, $\mu^*=\mu+b_1\sigma_\mu^2+d_1\sigma_e^2$, $\mathbf{V}_i=\sigma_\mu^2\mathbf{J}_{m_i}+\sigma_e^2\mathbf{I}_{m_i}$, $\mathbf{1}_{M_i}=(1,\ldots,1)'$ is a vector of length m_i, $\mathbf{J}_{m_i}$ is a square matrix of order m_i with every element equal one, and $\mathbf{I}_{m_i}$ is the identity matrix of order m_i.

Thus, the population distribution and the sample distribution of the cluster measurements, $\mathbf{y}_i=\left(y_{i1},\ldots,y_{im_i}\right)'$, belong to the same family, which is multivariate normal, but the mean in the sample is shifted by a constant, $b_1\sigma_\mu^2+d_1\sigma_e^2$, which is a function of the informativeness parameters b_1 and d_1.

Note that if the informativeness parameters $b_1=0$ and $d_1=0$, that is, the sampling design at the both stages is noninformative, then $\mu^*=\mu$, and hence the two-stage sample and population models are the same.

Equation (12) is used as a basis of maximum likelihood estimation of $\boldsymbol{\theta}=\left(\mu,\sigma_\mu^2,\sigma_e^2\right)$ under informative probability sampling design.

4. Fixed Effects Models for One-Way Classifications under Informative Sampling

4.1. *Single-Stage Population Model*

Consider the following single-stage population model that includes fixed effects:

$$y_{ij}=\mu_i+e_{ij};\quad i=1,\ldots,N;\quad j=1,\ldots,M_i$$

or

$$y_{ij}=\mu+\alpha_i+e_{ij};\quad i=1,\ldots,N;\quad j=1,\ldots,M_i;\quad \sum_{i=1}^{N}\alpha_i=0,$$

where e_{ij} are independent, with population distributions $N\left(0,\sigma_e^2\right)$.

In matrix notation this model can be expressed as:

$$\mathbf{y}_i=\mu_i\mathbf{1}_{M_i}+\mathbf{e}_i=\boldsymbol{\mu}_i+\mathbf{e}_i,$$
$$\boldsymbol{\mu}_i=\mu_i\mathbf{1}_{M_i},$$

$$E_p(\mathbf{e}_i) = \mathbf{0},$$
$$E_p(\mathbf{e}_i \mathbf{e}'_i) = Cov_p(\mathbf{e}_i) = \mathbf{V}_i = \sigma_e^2 \mathbf{I}_{M_i},$$

where $\mathbf{y}_i = \left(y_{i1}, \ldots, y_{iM_i}\right)'$, $\mathbf{1}_{M_i} = (1, \ldots, 1)'$ is a vector of length M_i, $\mathbf{I}_{M_i}$ is the identity matrix of order M_i, and the population distribution of $\mathbf{e}_i = \left(e_{i1}, \ldots, e_{iM_i}\right)'$ is $N(\mathbf{0}, \mathbf{V}_i)$. Hence, the population distribution of $\mathbf{y}_i$ is $N\left(\mu_i \mathbf{1}_{M_i}, \mathbf{V}_i\right)$. That is,

$$f_p(\mathbf{y}_i) = \frac{1}{(2\pi)^{M_i/2}} |\mathbf{V}_i|^{1/2} \exp\left\{-\frac{1}{2}\left(\mathbf{y}_i - \mu_i \mathbf{1}_{M_i}\right)' \mathbf{V}_i^{-1}\left(\mathbf{y}_i - \mu_i \mathbf{1}_{M_i}\right)\right\}$$
$$= \frac{1}{(2\pi)^{M_i/2}} \left(\sigma_e^2\right)^{M_i/2} \exp\left\{-\frac{1}{2}\sum_{j=1}^{M_i}\left(y_{ij} - \mu_i\right)^2\right\}.$$

4.2. *Sample Design and Sample Marginal Distribution*

Since the factor effects are fixed, the two-stage sample design is not relevant. Thus, we assume a single-stage informative sample design. The special case in which sampling is non informative can be considered trivially. Let z_{ij}; $i = 1, \ldots, N$; $j = 1, \ldots, M_i$ be the design variables (considered as random), used for the sample selection but not included in the working model under consideration. From each fixed effect factor of size M_i, a sample s_i of size m_i units is selected from the ith factor effect with inclusion probabilities: $\pi_{ij} = \Pr\left(j \in s_i \,|\, y_{ij}, z_{ij}\right)$ for all units $j = 1, 2, \ldots, M_i$ and all factor effects $i = 1, 2, \ldots, N$. Let $w_{ij} = 1/\pi_{ij}$ be the sampling weights for $i = 1, \ldots, N$; $j = 1, \ldots, M_i$. In the following, we use only the conditional expectation of the inclusion probabilities $E_p\left(\pi_{ij} \,|\, y_{ij}\right) = \exp\left(d_0 + d_1 y_{ij}\right)$. Using the results obtained in Sec. 4.1 with some modifications, we have:

(a) the conditional sample pdf of y_{ij} is given by:

$$f_s\left(y_{ij} \,|\, \mu_i\right) = \left(2\pi\sigma_e^2\right)^{-0.5} \exp\left(-\frac{1}{2\sigma_e^2}\left(y_{ij} - \left(\mu_i + d_1\sigma_e^2\right)\right)^2\right); \tag{13}$$

(b) the sample marginal distribution of the response variable measurements is:

$$f_s\left(y_{i1}, \ldots, y_{im_i} \,|\, \mu_i\right) = \prod_{j=1}^{m_i} f_s\left(y_{ij} \,|\, \mu_i\right)$$
$$= \left(2\pi\sigma_e^2\right)^{-0.5m_i} \exp\left(-\frac{1}{2\sigma_e^2}\sum_{j=1}^{m_i}\left(y_{ij} - \mu_i^*\right)^2\right), \tag{14}$$

where $\mu_i^* = \mu_i + d_1\sigma_e^2$.

(c) according to Eqs. (13) and (14), the fixed effects sample model is given by:

$$\mathbf{y}_i = \mu_i^* \mathbf{1}_{m_i} + \mathbf{e}_i = \boldsymbol{\mu}_i^* + \mathbf{e}_i,$$
$$\mu_i^* = \mu_i + d_1\sigma_e^2,$$
$$\boldsymbol{\mu}_i^* = \mu_i^* \mathbf{1}_{m_i} \tag{15}$$
$$E_s(\mathbf{e}_i) = \mathbf{0},\ E_s(\mathbf{e}_i \mathbf{e}'_i) = Cov_s(\mathbf{e}_i) = \mathbf{V}_i = \sigma_e^2 \mathbf{I}_{m_i},$$

where $\mathbf{1}_{m_i} = (1, \ldots, 1)'$ is a vector of length m_i and the sample distribution of $\mathbf{y}_i$ is $N(\boldsymbol{\mu}_i^*, \mathbf{V}_i)$; $i = 1, \ldots, N$.

5. Estimation of Population Parameters

5.1. *Variance Components Models under Informative Sampling*

Here, we are interested in estimating the vector of unknown population parameters, $\boldsymbol{\theta} = (\mu, \sigma_\mu^2, \sigma_e^2)$, that characterize the population variance components model given in Eqs. (4) and (5). We consider four different estimating methods: a two-step maximum likelihood method; an unweighted (exact) maximum likelihood (UWML) for the case where the sampling design is ignorable; a pseudo maximum likelihood (PML) method; and ANOVA estimation.

5.1.1. *Maximum Likelihood Estimation—Two-Step Method.* In this method, we base the inference on the sample pdf given in Eq. (12).

Step 1. I. Estimation of $E_p(\pi_i \mid \mu_i) = \exp(b_0 + b_1\mu_i)$: based on the relationship given in Eq. (2), we have, approximately:

$$\ln(w_i) = W_i = -b_0 - b_1\mu_i + k_i,$$

where k_i are uncorrelated random variables with $E(k_i) = 0$ and $Var(k_i) = \sigma_k^2$. Assuming $\bar{y}_i = \mu_i + h_i$, where $\bar{y}_{i\cdot} = m_i^{-1}\sum_{j=1}^{m_i} y_{ij}$, h_i are uncorrelated random variables with $E(h_i) = 0$ and $Var(h_i) = \sigma_h^2/m_i$. Then by the measurement error model of Fuller (1987, Sec. 1.2), we obtain:

$$\hat{b}_0 = -\left(\overline{W} - \hat{b}_1\bar{y}_{\cdot\cdot}\right),$$

where

$$\bar{y}_{\cdot\cdot} = m^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}, \quad m = \sum_{i=1}^{n} m_i$$

and

$$\overline{W} = n^{-1}\sum_{i=1}^{n} W_i.$$

Also, if σ_h^2 is known, then

$$\hat{b}_1 = -\left(m_{\bar{y}\bar{y}} - \frac{1}{n}\sum_{i=1}^{n}\frac{\sigma_h^2}{m_i}\right)^{-1} m_{\bar{y}W} \tag{16}$$

where

$$m_{\bar{y}\bar{y}} = (n-1)^{-1}\sum_{i=1}^{n}(\bar{y}_i - \bar{y}_{\cdot\cdot})^2,$$

and

$$m_{\bar{y}W} = (n-1)^{-1} \sum_{i=1}^{n} (\bar{y}_i - \bar{y}_{..}) \left(W_i - \overline{W}\right)^2.$$

If σ_h^2 is unknown we can estimate it by:

$$\hat{\sigma}_h^2 = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m_i} \left(y_{ij} - \bar{y}_i\right)^2}{\sum_{i=1}^{n} (m_i - 1)}.$$

II. Estimation of $E_p\left(\pi_{j?i} \mid y_{ij}, \mu_i\right) = \exp\left(d_0 + d_1 y_{ij}\right)$: based on the relationship given in Eq. (2), we have, approximately:

$$\ln\left(w_{j|i}\right) = -d_0 - d_1 y_{ij} + r_j,$$

where r_j are uncorrelated random variables with $E\left(r_j\right) = 0$ and $Var\left(r_j\right) = \sigma_r^2$. $i = 1, \ldots, n$; $j = 1, \ldots, m_i$. Thus, the least square estimator of $\mathbf{d} = (d_0, d_1)'$ is given by:

$$\hat{d} = \left(\hat{d}_0, \hat{d}_1\right)' = -\left(\mathbf{Y}'Y\right)^{-1} \mathbf{Y}'W \tag{17}$$

where

$$\mathbf{Y} = \begin{bmatrix} 1 & \ldots & 1 & \ldots & 1 & \ldots & 1 \\ y_{11} & \ldots & y_{1m_1} & \ldots & y_{n1} & \ldots & y_{nm_n} \end{bmatrix}'$$

$$\mathbf{W} = \left(\ln w_{1|1}, \ldots, \ln w_{m_i|1}, \ldots, \ln w_{1|n}, \ldots, \ln w_{m_n|n}\right)'$$

Step 2. Having estimated the informativeness parameters b_1 and d_1, we plug the estimates into the sample model given in Eq. (12) and then, in the next step, use this sample model to estimate the parameters of the population model, given in Eqs. (4) and (5). This is done by maximizing the following resulting sample log-likelihood function:

$$\begin{aligned} L_{rs}\left(\mu, \sigma_e^2, \sigma_\mu^2\right) &= \ln \prod_{i=1}^{n} f_s\left(y_{i1}, \ldots, y_{im_i}\right) = \sum_{i=1}^{n} \ln f_s\left(y_{i1}, \ldots, y_{im_i}\right) \\ &= \sum_{i=1}^{n} \left(-0.5\,(m_i - 1) \log\left(\sigma_e^2\right) - 0.5 \log\left(m_i \sigma_\mu^2 + \sigma_e^2\right)\right) \\ &\quad - \frac{1}{2\sigma_e^2} \sum_{i=1}^{n} \left(\sum_{j=1}^{m_i} \left(y_{ij} - \left(\mu + \hat{b}_1 \sigma_\mu^2 + \hat{d}_1 \sigma_e^2\right)\right)^2 \right) \\ &\quad + \sum_{i=1}^{n} \frac{\sigma_\mu^2}{2\left(\sigma_e^2\right)\left(m_i \sigma_\mu^2 + \sigma_e^2\right)} \left(\sum_{j=1}^{m_i} \left(y_{ij} - \left(\mu + \hat{b}_1 \sigma_\mu^2 + \hat{d}_1 \sigma_e^2\right)\right) \right)^2 \end{aligned} \tag{18a}$$

This function can be maximized by numerical methods (e.g., using the nlminb function in S-PLUS, Statistical Sciences, 1990).

For the variance estimation of $\hat{\boldsymbol{\theta}} = \left(\hat{\mu}, \hat{\sigma}_e^2, \hat{\sigma}_\mu^2\right)$, we consider the use of the inverse of Fisher information matrix, following Pfeffermann and Sverchkov (1999, 2003).

We first consider estimating the conditional variance of $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}_e^2, \hat{\sigma}_\mu^2)$, given that the informativeness parameters (b_1, d_1) are held fixed at their estimated values. The conditional Fisher information matrix evaluated at $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}_e^2, \hat{\sigma}_\mu^2)$ is given by:

$$\widehat{V}_{sample}\left(\hat{\boldsymbol{\theta}}\right) = \left[I_{sample}\left(\hat{\boldsymbol{\theta}}\right)\right]^{-1} = \left\{-\frac{1}{n}\left[\frac{\partial^2 L_{rs}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'\partial\boldsymbol{\theta}}\right]\Bigg|_{\theta=\hat{\theta}}\right\}^{-1}.$$

For instance, under the sample log-likelihood function given in (18a), the entries of $I_{sample}(\hat{\boldsymbol{\theta}})$ are easily computed.

In order to estimate the unconditional variance, the unconditional sample likelihoods must be used; see Eq. (18a) with $(\hat{b}_1, \hat{d}_1) = (b_1, d_1)$. An alternative to the Fisher information method that can be used is the bootstrap approach for variance estimation, where first the sampled psu's are selected with replacement and then final units are selected from these selected psu's with replacement. This is well founded under informative sampling, because as mentioned before, under many sampling schemes used in practice, such as successive sampling, rejective sampling, and Sampford's method, the sample measurements are asymptotically independent with respect to the sample distribution, see Pfeffermann et al. (1998a). Let $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}_e^2, \hat{\sigma}_\mu^2)$ be the sample MLE of $\boldsymbol{\theta} = (\mu, \sigma_e^2, \sigma_\mu^2)$ obtained based on Eq. (18a) and $\hat{\boldsymbol{\theta}}_q = (\hat{\beta}, \hat{\gamma}, \hat{\sigma}_e^2, \hat{\sigma}_\mu^2)$ be the ML estimator computed from the bootstrap sample $q = 1, \ldots, B$, with the same sample size, drawn by simple random sampling with replacement from the original sample – the sample drawn under informative sampling design. The bootstrap variance estimator of $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}_e^2, \hat{\sigma}_\mu^2)$ is defined as:

$$\widehat{V}_{boot}\left(\hat{\boldsymbol{\theta}}\right) = \frac{1}{B}\sum_{q=1}^{B}\left(\hat{\boldsymbol{\theta}}_q - \hat{\boldsymbol{\theta}}_{boot}\right)\left(\hat{\boldsymbol{\theta}}_q - \hat{\boldsymbol{\theta}}_{boot}\right)',$$

where

$$\hat{\boldsymbol{\theta}}_{boot} = \frac{1}{B}\sum_{q}^{B}\hat{\boldsymbol{\theta}}_q.$$

As pointed out by Pfeffermann and Sverchkov (2003), "a possible advantage of the use of bootstrap variance estimator in the present context is that it accounts in principle for all sources of variation, including that due to the estimation of the unknown informativeness parameters b_1 and d_1, so that it estimates the unconditional variance."

5.1.2. *Unweighted (exact) Maximum Likelihood Estimation.* The estimator of $\boldsymbol{\theta} = (\mu, \sigma_\mu^2, \sigma_e^2)$ for the case where the sampling design is ignorable can be obtained by setting $\hat{b}_1 = 0$ and $\hat{d}_1 = 0$ in Eq. (18a), and also by maximizing:

$$L_{srs}\left(\mu, \sigma_e^2, \sigma_\mu^2\right) = \sum_{i=1}^{n}\left(-0.5\,(m_i - 1)\log\left(\sigma_e^2\right) - 0.5\log\left(m_i\sigma_\mu^2 + \sigma_e^2\right)\right) - \frac{1}{2\sigma_e^2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(y_{ij} - \mu\right)^2 + \sum_{i=1}^{n}\frac{\sigma_\mu^2}{2\left(\sigma_e^2\right)\left(m_i\sigma_\mu^2 + \sigma_e^2\right)}\left(\sum_{j=1}^{m_i}\left(y_{ij} - \mu\right)\right)^2 \tag{18b}$$

with respect to $\boldsymbol{\theta} = (\mu, \sigma_\mu^2, \sigma_e^2)$.

5.1.3. *Pseudo Maximum Likelihood (PML) Estimation.* Next, we extend the idea of PML (Binder, 1983) to two-stage cluster sampling. According to Sec. 3, the first-stage inclusion probabilities are denoted by π_i; $i = 1, \ldots, N$ and the second-stage inclusion probabilities are denoted by $\pi_{j|i}$; $j = 1, \ldots, M_i$. So that the joint inclusion probabilities are given by $\pi_{ij} = \pi_i \pi_{j|i}$; $i = 1, \ldots, N$; $j = 1, \ldots, M_i$. Therefore, the joint sample weights are given by:

$$w_{ij} = w_i w_{j?i}, \quad w_i = \pi_i^{-1}, \quad w_{j|i} = \pi_{j|i}^{-1}; \quad i = 1, \ldots, N; \quad j = 1, \ldots, M_i.$$

Under the conditions of the two-stage population model given in Eqs. (4) and (5), and using Eq. (6), the census maximum likelihood estimator of $\boldsymbol{\theta} = \left(\mu, \sigma_\mu^2, \sigma_e^2\right)$ solves the census likelihood equations, which in our case are:

$$U(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{\partial L_{C_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0} = (0, 0, 0)'$$

where

$$\frac{\partial L_{C_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left(\frac{\partial L_{C_i}\left(\mu, \sigma_\mu^2, \sigma_e^2\right)}{\partial \mu}, \frac{\partial L_{C_i}\left(\mu, \sigma_\mu^2, \sigma_e^2\right)}{\partial \sigma_\mu^2}, \frac{\partial L_{C_i}\left(\mu, \sigma_\mu^2, \sigma_e^2\right)}{\partial \sigma_e^2} \right)'$$

and

$$L_{C_i}(\boldsymbol{\theta}) = \ln f_p(\mathbf{y}_i) = -0.5(M_i - 1)\log\left(\sigma_e^2\right) - 0.5\log\left(M_i\sigma_\mu^2 + \sigma_e^2\right) - \frac{1}{2\sigma_e^2}\sum_{j=1}^{M_i}\left(y_{ij} - \mu\right)^2 + \frac{\sigma_\mu^2}{2\left(\sigma_e^2\right)\left(M_i\sigma_\mu^2 + \sigma_e^2\right)}\left(\sum_{j=1}^{M_i}\left(y_{ij} - \mu\right)\right)^2.$$

The pseudo maximal likelihood (PML) estimator is defined as the solution of $\widehat{U}(\boldsymbol{\theta}) = 0$ where $\widehat{U}(\boldsymbol{\theta})$ is a sample estimator of the census log-likelihood, $U(\boldsymbol{\theta})$.

Now the probability weighted estimator of $L_{C_i}\left(\boldsymbol{\theta}_p\right)$ is given by:

$$\widehat{L}_{C_i}(\boldsymbol{\theta}) = -0.5\left(\sum_{j=1}^{m_i} w_{j|i} - 1\right)\log\left(\sigma_e^2\right) - 0.5\log\left(\left(\sum_{j=1}^{m_i} w_{j|i}\right)\sigma_\mu^2 + \sigma_e^2\right) - \frac{1}{2\sigma_e^2}\sum_{j=1}^{m_i} w_{j|i}\left(y_{ij} - (\mu)\right)^2 + \frac{\sigma_\mu^2}{2\left(\sigma_e^2\right)\left(\left(\sum_{j=1}^{m_i} w_{j|i}\right)\sigma_\mu^2 + \sigma_e^2\right)}\left(\sum_{j=1}^{m_i} w_{j|i}\left(y_{ij} - (\mu)\right)\right)^2, \tag{18c}$$

where $\sum_{j=1}^{n} w_{j|i}$ is an unbiased estimator of M_i. Thus, the PML estimator is defined as the solution of following estimating equation:

$$\widehat{U}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} w_{ij}\frac{\partial \widehat{L}_{C_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} w_i \frac{\partial \widehat{L}_{C_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}.$$

5.1.4. *Analysis of Variance (ANOVA) Estimation.* The ANOVA estimators of σ_e^2 and σ_μ^2 are based on the expected sum of squares under the sample model given in Eq. (11). The analysis of variance sums of squares, for the unbalanced sample variance components model given in Eq. (11), based on the sample data are:

$$\begin{aligned} SST &= \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(y_{ij}-\bar{y}_{..}\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(y_{ij}\right)^2 - Ny_{..}^2 \\ SSE &= \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(y_{ij}-\bar{y}_{i.}\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(y_{ij}\right)^2 - \sum_{i=1}^{n} m_i^{-1} y_{i.}^2, \\ SSA &= \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\bar{y}_{i.}-\bar{y}_{..}\right)^2 = \sum_{i=1}^{n} m_i^{-1} y_{i.}^2 - Ny_{..}^2, \\ SST &= SSA + SSE. \end{aligned} \tag{19}$$

Expected Sums of Squares under the Sample Model given in Eq. (11). Under the sample model given in Eq. (11), we have the following expectations with respect to the sample distributions:

$$\begin{aligned} E_s\left(y_{ij}\right) &= \mu + b_1\sigma_\mu^2 + d_1\sigma_e^2 \\ E_s\left(y_{ij}^2\right) &= \left(\sigma_\mu^2 + \sigma_e^2\right) + \left(\mu + b_1\sigma_\mu^2 + d_1\sigma_e^2\right)^2 \\ E_s\left(\bar{y}_{i.}\right) &= \mu + b_1\sigma_\mu^2 + d_1\sigma_e^2 \\ E_s\left(\bar{y}_{i.}^2\right) &= m_i^{-1}\sigma_e^2 + \sigma_\mu^2 + \left(\mu + b_1\sigma_\mu^2 + d_1\sigma_e^2\right)^2. \end{aligned}$$

Also, we can show that:

$$\begin{aligned} E_s\left(SSE\right) &= \sum_{i=1}^{n}\left(E_s\left(\sum_{j=1}^{m_i} y_{ij}^2\right) - m_i E_s\left(\bar{y}_{i.}^2\right)\right) \\ &= \sum_{i=1}^{n}\left(m_i - 1\right)\sigma_e^2 \\ &= \left(m-n\right)\sigma_e^2. \end{aligned}$$

It can be shown that the formula for $E_s\left(SSA\right)$ is given by:

$$\begin{aligned} E_s\left(SSA\right) &= E_s\left(\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\bar{y}_{i.}-\bar{y}_{..}\right)^2\right) \\ &= \sum_{i=1}^{n}\left(m_i E_s\left(\bar{y}_{i.}^2\right)\right) - E_s\left(N\bar{y}_{..}\right) \\ &= \left(\left(n-1\right)\sigma_e^2\right) + \sigma_\mu^2\left(m - m^{-1}\sum_{i=1}^{n} m_i^2\right). \end{aligned}$$

ANOVA Estimators of σ_e^2 and σ_μ^2. Having derived $E_s(SSE)$ and $E_s(SSA)$ we use these expressions to equate sums of squares (or, equivalently, mean squares) to their

expected values—the ANOVA method of estimation. The equations are:

$$SSE = (m-n)\,\hat{\sigma}_e^2$$
$$SSA = (n-1)\,\hat{\sigma}_e^2 + \left(m - m^{-1}\sum_{i=1}^{n} m_i^2\right)\hat{\sigma}_\mu^2.$$

These yield the ANOVA estimators:

$$\hat{\sigma}_e^2 = \frac{SSE}{(m-n)} = MSE, \tag{20a}$$

and

$$\hat{\sigma}_\mu^2 = \frac{SSA - (n-1)\,\hat{\sigma}_e^2}{\left(m - m^{-1}\sum_{i=1}^{n} m_i^2\right)} = \frac{(n-1)\,(MSA - MSE)}{\left(m - m^{-1}\sum_{i=1}^{n} m_i^2\right)}, \tag{20b}$$

where $MSA = SSA/n - 1$.

Note that these ANOVA estimators are the estimators obtained for the variance components model under noninformative sample design. This is intuitively reasonable, because under the exponential conditional expectation of first order inclusion probabilities, see models given in Eqs. (7) and (9), and when the population distribution is normally distributed, the sample pdf is also normal with same variance and different mean. That is, informativeness impacts on the mean but not on the variances. Therefore, as far as the variance components are concerned and we obtained for ANOVA-type estimators, we can overlook informativeness. But if the conditional expectation of first-order inclusion probabilities is not of exponential type, the mean and the variance are changing. For more information on the effect of modeling $E_p(\pi_i \mid y_i, \gamma)$ on the sample models, see Eideh (2010).

Estimating the Mean μ. According to Eq. (11), our sample model can be written in matrix form as:

$$\mathbf{y}_i = \mathbf{1}_{m_i}\mu^* + \mathbf{u}_i,\ E_s(\mathbf{u}_i) = \mathbf{0},$$
$$\mu^* = \mu + b_1\sigma_\mu^2 + d_1\sigma_e^2$$
$$E_s(\mathbf{u}_i\mathbf{u}'_i) = Cov_s(\mathbf{u}_i) = \mathbf{V}_i = \sigma_\mu^2\mathbf{J}_{m_i} + \sigma_e^2\mathbf{I}_{m_i},$$

where $\mathbf{1}_{m_i} = (1, \ldots, 1)'$ is a vector of length m_i and the sample distribution of $\mathbf{y}_i$ is $N\left(\mathbf{1}_{m_i}\mu^*, \mathbf{V}_i\right)$

Assuming $\mathbf{V}_i$ is known, the generalized least squares estimate of μ^* is the value $\hat{\mu}^*$ which minimizes the quadratic form:

$$q(\mu^*) = \sum_{i=1}^{n}\left(\mathbf{y}_i - \mathbf{1}_{m_i}\mu^*\right)'\mathbf{V}_i^{-1}\left(\mathbf{y}_i - \mathbf{1}_{m_i}\mu^*\right).$$

Standard matrix manipulations give the explicit result:

$$\hat{\mu}^* = \left[\sum_{i=1}^{n}\left(\mathbf{1}'_{m_i}\mathbf{V}_i^{-1}\mathbf{1}_{m_i}\right)\right]^{-1}\left[\sum_{i=1}^{n}\mathbf{1}'_{m_i}\mathbf{V}_i^{-1}\mathbf{y}_i\right],$$
$$= \left[\sum_{i=1}^{n} v_i\right]^{-1}\left[\sum_{i=1}^{n} v_i\bar{y}_{i.}\right]$$

where

$$v_i = \frac{m_i}{m_i\sigma_\mu^2 + \sigma_e^2} = \left(\sigma_\mu^2 + \frac{\sigma_e^2}{m_i}\right)^{-1} = \left(Var_s\left(\bar{y}_{i.}\right)\right)^{-1}.$$

Note that $\hat{\mu}^*$ is a weighted average of the cluster means $\{\bar{y}_{i.}, i = 1, \ldots, n\}$ with weights $\{v_i, i = 1, \ldots, n\}$. If σ_e^2 and $\sigma_\mu^2$are unknown, we replace them by their ANOVA estimators or by their maximum likelihood estimators (see below). Thus the generalized least squares estimator of μ is the solution of the equation:

$$\hat{\mu}^* = \hat{\mu} + \hat{b}_1\hat{\sigma}_\mu^2 + \hat{d}_1\hat{\sigma}_e^2,$$

which is given by:

$$\hat{\mu} = \hat{\mu}^* - \hat{b}_1\hat{\sigma}_\mu^2 - \hat{d}_1\hat{\sigma}_e^2 = \left(\sum_{i=1}^{n} v_i\right)^{-1}\left(\sum_{i=1}^{n} v_i\bar{y}_{i.}\right) - \hat{b}_1\hat{\sigma}_\mu^2 - \hat{d}_1\hat{\sigma}_e^2, \tag{21}$$

where $\hat{b}_1$ and $\hat{d}_1$ are given in Eqs. (16) and (17), respectively. If $b_1 = d_1 = 0$, that is the sample design for both the two stages is noninformative, then $\hat{\mu} = \hat{\mu}^*$, which is the classical generalized least squares estimator of μ, obtained under a noninformative sampling mechanism. Also note that the generalized least squares estimator $\hat{\mu}^*$ is the maximum likelihood estimator under the multivariate normal assumption.

As clearly indicated by Eq. (21), in particular, if $\hat{b}_1 > 0$ and $\hat{d}_1 > 0$, then the generalized least squares estimator of μ is smaller than the generalized least squares estimator of μ under noninformative sampling. Thus, the use of the generalized least squares estimator of μ that ignores the sampling process yields a biased estimator in this case.

5.2. *Fixed Effects Models for One-Way Classification under Informative Sampling*

Here, we are interested in estimating μ_i and σ_e^2. Similar to Sec. 5.1, we consider four different estimators.

(a) The sample log-likelihood estimators of $\left(\mu_i, \sigma_e^2\right)$ are obtained by maximizing:

$$L_{rs}\left(\mu_i, \sigma_e^2\right) = \sum_{i=1}^{N} -0.5\left(m_i - 1\right)\log\left(\sigma_e^2\right) - \frac{1}{2\sigma_e^2}\sum_{j=1}^{m_i}\left(y_{ij} - \left(\mu_i + \hat{d}_1\sigma_e^2\right)\right)^2, \tag{22}$$

where $\hat{d}_1$ is obtained from (14): $\hat{d} = (\hat{d}_0, \hat{d}_1)' = -(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{W}$ but

$$\mathbf{Y} = \begin{bmatrix} 1 & \ldots & 1 & \ldots & 1 & \ldots & 1 \\ y_{11} & \cdots & y_{1m_1} & \cdots & y_{N1} & \cdots & y_{Nm_n} \end{bmatrix}'$$

and

$$\mathbf{W} = \left(\ln w_{11}, \ldots, \ln w_{1m_1}, \ldots, \ln w_{N1}, \ldots, \ln w_{Nm_N}\right)'.$$

(b) The unweighted maximum likelihood estimator of $\left(\mu_i, \sigma_e^2\right)$ for the case where the sampling design is ignorable can obtained by setting $\hat{d}_1 = 0$ in Eq. (22), and maximizing:

$$L_{srs}\left(\mu_i, \sigma_e^2\right) = \sum_{i=1}^{N} -0.5\left(m_i - 1\right)\log\left(\sigma_e^2\right) - \frac{1}{2\sigma_e^2}\sum_{j=1}^{m_i}\left(y_{ij} - \mu_i\right)^2$$

with respect to $\left(\mu_i, \sigma_e^2\right)$.

(c) The PML estimator of $\left(\mu_i, \sigma_e^2\right)$ is defined as the solution of the following estimating equations:

$$\frac{\partial \widehat{L}_{CH}\left(\mu_i, \sigma_e^2\right)}{\partial \mu_i} = \sum_{i=1}^{N}\frac{\partial \widehat{L}_{CH_i}\left(\mu_i, \sigma_e^2\right)}{\partial \mu_i} = 0$$

$$\frac{\partial \widehat{L}_{CH}\left(\mu_i, \sigma_e^2\right)}{\partial \sigma_e^2} = \sum_{i=1}^{N}\frac{\partial \widehat{L}_{CH_i}\left(\mu_i, \sigma_e^2\right)}{\partial \sigma_e^2} = 0,$$

where

$$\widehat{L}_{CH_i}\left(\mu_i, \sigma_e^2\right) = -0.5\left(\sum_{j=1}^{m_i} w_{ij} - 1\right)\log\left(\sigma_e^2\right) - \frac{1}{2\sigma_e^2}\sum_{j=1}^{m_i} w_{ij}\left(y_{ij} - \mu_i\right)^2$$

and $\sum_{j=1}^{m_i} w_{ij}$ is an unbiased estimator of M_i.

(d) The analysis of variance sums of squares for unbalanced fixed effects sample model, see Eq. (15), based on the sample data are similar to those of Eq. (19), but with n replaced by N. Then, we can show that:

$$E_s\left(SSE\right) = \sum_{i=1}^{N}\left(E_s\left(\sum_{j=1}^{m_i} y_{ij}^2\right) - m_i E_s\left(\bar{y}_{i.}^2\right)\right)$$

$$= \sum_{i=1}^{N}\left(m_i - 1\right)\sigma_e^2 = \left(m - N\right)\sigma_e^2.$$

This yields the ANOVA estimator of σ_e^2:

$$\hat{\sigma}_e^2 = \frac{SSE}{(m - N)} = MSE$$

Estimating the Means μ_i. Under the sample model given in Eq. (15), assuming σ_e^2 is known, we can show that the generalized least squares estimate of μ_i^* is:

$$\hat{\mu}_i^* = \hat{\mu}_i + \hat{d}_1\sigma_e^2 = \left[\sum_{i=1}^{N}\frac{m_i}{\sigma_e^2}\right]^{-1}\left[\sum_{i=1}^{N}\frac{m_i}{\sigma_e^2}\bar{y}_{i.}\right] = \sum_{i=1}^{N}\frac{m_i}{m}\bar{y}_{i.},$$

where $m = \sum_{i=1}^{N} m_i$.

If σ_e^2 is unknown, we replace it by one of the four estimators (see above). Thus the generalized least squares estimator of μ_i is the solution of the equation: $\hat{\mu}_i^* = \hat{\mu}_i + \hat{d}_1\hat{\sigma}_e^2$, which is given by:

$$\hat{\mu}_i = \hat{\mu}_i^* - \hat{d}_1\hat{\sigma}_e^2 = \sum_{i=1}^{N} \frac{m_i}{m}\bar{y}_{i.} - \hat{d}_1\hat{\sigma}_e^2.$$

In particular, if $d_1 = 0$, that is the sample design is non informative, then $\hat{\mu}_i = \hat{\mu}_i^*$, which is the classical generalized least squares estimator of $\hat{\mu}_i$, obtained under a noninformative sampling mechanism.

6. Simulation Study

In order to assess the performance of the estimators obtained using sample likelihood, analysis of variance, and pseudo likelihood procedures under informative sampling. Then we compare them with the classical estimators obtained under the assumption of ignorability of the sampling design, a simulation study was carried out.

6.1. *Generation of Population Values*

The population values were generated in two steps.

Step 1. We generated independently univariate normal values of the primary sampling unit-specific-effects, μ_i, of size $N = 10{,}000$ from: $\mu_i \sim N\left(\mu, \sigma_\mu^2\right)$, where $i = 1, \ldots, N$, $\mu = 1$ and $\sigma_\mu^2 = 0.36$.

Step 2. We generated independently the population values of the secondary sampling units from: $y_{ij} = \mu + \eta_i + e_{ij}$; $j = 1, \ldots 100$, where $e_{ij} \sim N(0, 0.64)$.

6.2. *Sample Selection*

Single-stage samples of size $n = 100$ primary sampling units were selected by probability proportional to size systematic sampling, with the size variable defined by the exponential sampling model: $z_i = \exp(1.2 + 0.9\mu_i)$. Under this sampling scheme, the first stage inclusion probabilities are defined by:

$$\pi_i = 100\frac{z_i}{Z} \quad \text{where } Z = \sum_{i=1}^{10000} z_i.$$

We assume that N is sufficiently large to ensure that π_i will not in practice exceed one.

The population was simulated $R = 10,000$ times and for each simulated population, samples of primary sampling units were independently drawn using probability proportional to size systematic sampling. Data from these samples were then used to estimate the informativeness parameters and then the population parameters using the exact ML, sample ML, pseudo ML, and the ANOVA procedures described in Sec. 5.

Table 1
Relative biases (RB) and relative root mean square errors (RRMSE) of four estimation methods

Parameter	Indicators	ANOVA	ML	PML	SML
μ	RB	0.00034	0.00074	0.00122	0.00032
	RRMSE	0.01247	0.01181	0.01254	0.01243
σ_μ^2	RB	0.05222	0.05118	0.04990	0.05201
	RRMSE	0.1501	0.15441	0.17280	0.1511
σ_e^2	RB	0.00311	0.00320	0.00375	0.00301
	RRMSE	0.01658	0.01669	0.01854	0.01637

6.3. *Results of the Simulation Study*

Now we report and discuss the results obtained for the simulation study described above. The parameters estimated in our study are the components of the vector $\boldsymbol{\theta} = \left(\mu, \sigma_\mu^2, \sigma_e^2\right)$. We consider four different estimators. These estimators are described as follows:

ML—Unweighted (exact) maximum likelihood (ML) for the case where the sampling design is ignorable and the estimators are obtained by maximizing Eq. (18b);
PML—The pseudo ML estimator obtained by maximizing Eq. (18c);
SML–The estimator, based on the sample distribution, obtained by maximizing Eq. (18a);
ANOVA—Analysis of variance estimators given in Eqs. (20a), (20b), and (21).

It should be noted here that, the likelihood functions were maximized using the *nlminb* function within S-Plus (Statistical Sciences, 1990).

The results of the simulation study are summarized in Table 1 as averages over the 10,000 samples selected under the exponential sampling scheme.

The relative bias of $\hat{\boldsymbol{\theta}}$ is estimated by:

$$RB\left(\hat{\boldsymbol{\theta}}\right) = \frac{1}{\boldsymbol{\theta}}\left(\frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}\right)\right). \tag{23}$$

The relative root mean square error of $\hat{\boldsymbol{\theta}}$ is estimated by:

$$RRMSE\left(\hat{\boldsymbol{\theta}}\right) = \frac{1}{\boldsymbol{\theta}}\left[\frac{1}{10000}\sum_{i=1}^{10000}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}\right)^2\right]^{0.5}. \tag{24}$$

Examination of the results in Table 1 shows the following.

1. The ML and PML estimators of μ are slightly biased. The ANOVA and sample maximum likelihood (SML), reduce this bias substantially. This result reflects the effect of selection bias, because in our case—exponential sampling—the mean under the sample model is different from the mean under the population model, see Eqs. (12) and (18).
2. The ML estimator of μ has the smallest RRMSE, while the PML and SML estimators of μ have the same RRMSE.

3. The RB and RRMSE of the estimators of σ_μ^2 and σ_e^2 are the same under the ML, ANOVA, and SML methods. This is because under exponential sampling the variances and covariances of measurements within primary sampling units do not change, see Eq. (12).
4. The PML estimators of σ_μ^2 and σ_e^2 have higher RRMSE. This RRMSE is reduced substantially by the other methods of estimation—ML, ANOVA, and SML.
5. The small differences between the performance of the ML, ANOVA, and SML estimation indicate that the effects of informativeness are very small in this experiment. This is probably due to the small variation in the first-stage selection probabilities – $S_\pi^2 = 0.0004$ and small correlation between μ_i and π_i, $Corr(\mu_i, \pi_i) = 0.091$.

In this article we considered only the exponential model for the conditional expectations of the inclusion probabilities for both stages. Other models can be used- see the references before Eq. (7). Eideh (2003) and Eideh and Nathan (2006a) showed that in many situations the sample likelihood method is not sensitive to the modeling of the conditional expectations of the inclusion probabilities.

7. Conclusions

In recent years there is a growth in the demand for fitting statistical models to complex survey data. In this article we fit the variance components model and fixed effects models to complex survey data, taking into account unequal probabilities of selection and informative sample designs. Also, we considered a new method of estimating the parameters of the two-stage and single-stage population models for two-stage and single-stage sampling from a finite population, when the sample design for the different stages is informative. Also, we extended the pseudo maximum likelihood estimation to the two-stage population model.

However, the main feature of the estimators presented in this article is their behaviour in terms of the informativeness parameters. Also, the use of the classical analysis of variance estimator or classical maximum likelihood estimator of the population mean obtained under the assumption of ignorability of the sample design yields biased estimators. Moreover, when the researcher does not have access to the design variables or decides not to include them in the modelling process, sample-based likelihood method of estimation is produce better estimators than other method considered in this article.

One of the advantages of the proposed approach is that, when weighted estimators are avoided, it is possible to study the finite sample distribution of the estimators.

Eventhough, the article is mostly mathematical, yet the role of informativeness of sampling mechanism in adjusting various estimators for bias reduction, can be found in Pfeffermann and Sverchkov (1999, 2003), and Eideh and Nathan (2006a,b, 2009) and Eideh (2008). This simulation approach was based on different population models and different modeling of conditional expectations of first order inclusion probabilities, given the values of the response variable and of the covariates. In particular, the properties of variance estimators based on Fisher information for fitting generalized linear models under informative sampling can be found in Pfeffermann and Sverchkov (2003).

We are certain that this new mathematical results obtained will encourage further theoretical, empirical and practical research in these directions.

Acknowledgements

The author is grateful to Gad Nathan, to the Associate Editor and to the referees for constructive comments and suggestions that helped improving the quality of the article.

References

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *Int. Statist. Rev.* 51:279–292.

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

Eideh, A. H. (2010). Analytic inference of complex survey data under informative probability sampling. In: Amin, Z., Hadi, A. S., eds. *Proc. Tenth Islamic Countries Confer. Statist. Sci. (ICCS-X)*. Vol. I. The Islamic Countries Society of Statistical Sciences, Lahore: Pakistan, Cairo: The American University, pp. 507–536.

Eideh, A. H. (2009). On the use of the sample distribution and sample likelihood for inference under informative probability sampling. *DIRASAT (Natural Science)* 36(1): 18–29.

Eideh, A. H., Nathan, G. (2009). Two-Stage informative cluster sampling with application in small area estimation. *J. Statist. Plann. Infer.* 139:3088–3101.

Eideh, A. H. (2008). Estimation and prediction of random effects models for longitudinal survey data under informative sampling. *Statist. Trans. New Series, 9, 3*, December pp. 485–502.

Eideh, A. H., Nathan, G. (2006a). The analysis of data from sample surveys under informative sampling. *Acta et Commentationes Universitatis Tartuensis de Mathematica*. Tartu 2006, 10:41–51.

Eideh, A. H., Nathan, G. (2006b). Fitting time series models for longitudinal survey data under informative sampling. *J. Statist. Plann. Infer.* 136(9):3052–306. [Corrigendum, 137 (2007), p. 628].

Eideh, A. H. (2003). Estimation for longitudinal survey data under informative sampling, PhD Thesis, Department of Statistics, Hebrew University of Jerusalem, Jerusalem, Israel.

Fuller, W. A. (1987). *Measurement Error Models*. New York: John Wiley.

Jia, Y., Stokes, L., Harris, I., Wang, Y. (2011). Performance of random effects model estimators under complex sampling designs. *J. Educat. Behav. Statist.* 36:6–32.

Korn, E. L., Graubard, B. I. (1999). *Analysis of Health Surveys*. New York: John Wiley.

Korn, E. L., Graubard, B. I. (2003). Estimating variance components by using survey data. *J. Roy. Statist. Soc. Ser. B* 1:175–190.

Nathan, G., Eideh, A. H. (2004). L'analyse des données issues des enquêtes longitudinales sous un plan de sondage informatif. In: Ardilly, P. *Échantillonage et Méthodes d'Enquêtes*. Paris: Dunod, pp. 227–240.

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

Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., Rasbash, J. (1998b). Weighting for unequal selection probabilities in multilevel models. *Journal. Roy. Statist. Soc. Ser. B* 60:23–40.

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

Pfeffermann, D., Sverchkov, M. (2003). Fitting generalized linear models under informative probability sampling. In: Chambers, R., Skinner, C. J., eds. *Analysis of Survey Data*. New York: Wiley, pp. 175–195.

Pfeffermann, D., Moura, F., Silva, P. (2006). Multi-level modeling under informative probability sampling. *Biometrika* 93:943–959.

Scott, A. J., Smith, T. M. F. (1969). Estimation in multistage surveys. *J. Amer. Statist. Assoc.* 64:830–840.

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

Skinner, C. J. (1994). Sample models and weights. *Amer. Statist. Assoc. Proc. Sec. Surv. Res. Meth.* 133–142.

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