Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 52 (2008) 2199–2217
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Performance of balanced two-stage empirical predictors of realized cluster latent values from finite populations: A simulation study

Silvina San Martino[a], Julio M. Singer[b,*], Edward J. Stanek III[c]

[a] *Facultad de Ciencias Agrarias, Universidad Nacional de Mar del Plata, Mar del Plata, Argentina*
[b] *Departamento de Estatística, Universidade de São Paulo, São Paulo, Brazil*
[c] *Department of Public Health, University of Massachusetts, Amherst, MA, USA*

Received 24 January 2007; received in revised form 16 July 2007; accepted 19 July 2007
Available online 31 July 2007

Abstract

Predictors of random effects are usually based on the popular mixed effects (*ME*) model developed under the assumption that the sample is obtained from a conceptual infinite population; such predictors are employed even when the actual population is finite. Two alternatives that incorporate the finite nature of the population are obtained from the superpopulation model proposed by Scott and Smith (1969. Estimation in multi-stage surveys. J. Amer. Statist. Assoc. 64, 830–840) or from the finite population mixed model recently proposed by Stanek and Singer (2004. Predicting random effects from finite population clustered samples with response error. J. Amer. Statist. Assoc. 99, 1119–1130). Predictors derived under the latter model with the additional assumptions that all variance components are known and that within-cluster variances are equal have smaller mean squared error (*MSE*) than the competitors based on either the *ME* or Scott and Smith's models. As population variances are rarely known, we propose method of moment estimators to obtain empirical predictors and conduct a simulation study to evaluate their performance. The results suggest that the finite population mixed model empirical predictor is more stable than its competitors since, in terms of *MSE*, it is either the best or the second best and when second best, its performance lies within acceptable limits. When both cluster and unit intra-class correlation coefficients are very high (e.g., 0.95 or more), the performance of the empirical predictors derived under the three models is similar.

Keywords: Empirical predictors; Finite population; Optimal prediction; Random permutation; Two-stage sampling; Mixed model; Superpopulation; Cluster sampling; Hierarchical models

1. Introduction

There are many instances where clustered finite populations occur naturally as in educational, public health or sociological surveys, where classrooms in schools, physician practices in hospitals or families in communities are typical examples of such clusters. In such settings, statistical inference is usually based on a multi-stage random sample selected without replacement. In addition to the sample average, three approaches may be considered to predict latent values of realized clusters (i.e., the average expected response of the units in those clusters). In each case, best linear

* Corresponding author. Tel.: +55 11 30916118; fax: +55 11 30916130.
E-mail addresses: ssanmarti@balcarce.inta.gov.ar (S. San Martino), jmsinger@ime.usp.br (J.M. Singer), stanek@schoolph.umass.edu (E.J. Stanek).

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.07.013

unbiased predictors (BLUP) have been obtained. The most popular approach is based on the usual mixed effects (*ME*) model derived under the assumption that the sample is obtained from a conceptual infinite population and is considered in Goldberger (1962), Henderson (1984), Kackar and Harville (1984), Prasad and Rao (1990), McLean et al. (1991), Robinson (1991), Stanek et al. (1999), Moura and Holt (1999) or McCulloch and Searle (2001) among others. The second approach, suggested by Scott and Smith (1969) and extended by Bolfarine and Zacks (1992) to include response error, considers the finite nature of the population and bases the inference on a superpopulation (*SP*) model. It has limited application, in part because its performance may be affected by model miss-specification. The third approach, recently suggested by Stanek et al. (2004) in a simple random sampling setup and extended by Stanek and Singer (2004) to a balanced two-stage sampling setup with or without response error, considers a finite population mixed (*FM*) model,[1] induced by the sampling process. Since the stochastic model is developed directly from two-stage sampling from a finite population, it can be applied to a wide range of practical settings.

In each case, the BLUPs of realized cluster latent values involve predictors of the response of the unobserved units and depend on weights called shrinkage factors; these shrinkage factors are functions of population variance components and of the number of sampled clusters. For the last two models, they also depend on finite population characteristics such as cluster sizes. The predictors obtained under these three models can occasionally be quite similar, but sometimes they can differ greatly.

As an example, suppose that an educational survey is conducted in a given high-school to evaluate the ability of second graders in a certain subject by means of a test with scores ranging from 0 to 10. We assume that the student responses include response error. To account for teacher effects, a two-stage random sample is obtained from the population of second grade students assigned to classrooms (each with 30 students). Assume that a sample of 15 students is selected from each classroom in a sample of classrooms in the school. In addition to estimating the school response and variance components, suppose that there is interest in predicting the latent response for a sampled classroom. For illustration, let us assume that the between classroom variance is 1.25, the within-cluster variance is 2.00 and that the response error variance is 0.80. With these assumptions, the cluster intra-class correlation coefficient is 0.38 and the unit intra-class correlation coefficient is 0.71 (see Section 2.4 for definitions of cluster and unit intra-class correlation coefficients). Also, suppose that the school sample average is 6.75, while for the ith sample classroom, the sample average is 5.20. Based on the sample data, there are four approaches to predict the latent classroom response. First, we may use the sample classroom average that is 5.20. Alternatively, assuming that the response error model holds for all students, the latent response for the ith sample classroom is predicted to be 5.40, 5.30 and 5.90, respectively, using the *ME*, *SP* or *FM* model predictors (see Section 2.4 for details about each predictor). The 11% relative difference observed between the predicted values obtained under the *FM* model and the *SP* model may be meaningful in this type of study. Consequently, an evaluation of the performance of the predictors derived under these three models for a wide range of conditions may be very helpful for practical applications. We consider such a comparison with the objective of selecting the predictor with smaller mean squared error (*MSE*).

The *ME*, *SP* and *FM* models can all be defined via a set of assumptions on the mean and on the covariance structure and do not require the specification of the form of the underlying distribution. Only the *FM* model links the finite population to the assumptions for the set of random variables that represent two-stage sampling (plus response error). When all variances are known and within-cluster variances are equal, Stanek and Singer (2004) show that the predictors of realized cluster latent values based on such a model have smaller *MSE* than those based on the other approaches. In practical situations, variances are rarely known and need to be estimated. We propose estimators for such variances and report simulation study results that compare the performance of empirical predictors of realized cluster latent values, providing guidance for the choice among the alternatives.

In Section 2 we present a brief review of the models and specify the corresponding predictors of sampled cluster latent values. We also propose empirical predictors based on variance components estimated from the sample. In Section 3 we describe technical details of the simulation study to compare the performance of these predictors for finite populations with different structures. Finally, in Sections 4 and 5 we present the simulation results and discussion, respectively. Programs and additional results are available at http://www.umass.edu/cluster/ed/.

[1] Although Stanek and Singer (2004) use the term "random permutation model", we prefer "finite population mixed model" in order to avoid confusion with the SP random permutation model of Hedayat and Sinha (1991).

2. BLUPs of the cluster latent value under different models

We consider a finite population with M units, indexed by $t = 1, \ldots, M$ in each of N clusters, indexed by $s = 1, \ldots, N$. A fixed constant y_{st}, called a "unit parameter", is associated with unit t in cluster s. We summarize these parameters in the vector

$$\mathbf{y} = (\mathbf{y}_1' \quad \mathbf{y}_2' \quad \cdots \quad \mathbf{y}_N')' \quad \text{where } \mathbf{y}_s = (y_{s1} \quad y_{s2} \quad \cdots \quad y_{sM})', \ s = 1, \ldots, N.$$

For $s = 1, \ldots, N$, we define the latent value of cluster s as $\mu_s = (1/M)\sum_{t=1}^{M} y_{st}$ and the corresponding variance as

$$\left(\frac{M-1}{M}\right)\sigma_s^2 = \frac{1}{M}\sum_{t=1}^{M}(y_{st} - \mu_s)^2.$$

Also, we let $\sigma_e^2 = (1/N)\sum_{s=1}^{N}\sigma_s^2$ denote the average within-cluster variance. Similarly, we define the population mean and the between-cluster variance as $\mu = (1/N)\sum_{s=1}^{N}\mu_s$ and

$$\left(\frac{N-1}{N}\right)\sigma^2 = \frac{1}{N}\sum_{s=1}^{N}(\mu_s - \mu)^2,$$

respectively. Letting $\beta_s = (\mu_s - \mu)$ denote the deviation of the latent value of cluster s from the population mean and $\varepsilon_{st} = (y_{st} - \mu_s)$ denote the deviation of the parameter for unit t (in cluster s) from the latent value of cluster s, we can re-parameterize the vector of fixed values $\mathbf{y}$ via the non-stochastic model

$$\mathbf{y} = \mathbf{X}\mu + \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where $\mathbf{X} = 1_{NM}$, $\mathbf{Z} = \mathbf{I}_N \otimes \mathbf{1}_M$, $\boldsymbol{\beta}' = (\beta_1 \ \beta_2 \ \cdots \ \beta_N)$ and $\boldsymbol{\varepsilon}$ is defined similarly to $\mathbf{y}$. Here, $\mathbf{1}_a$ denotes an $a \times 1$ column vector with all elements equal to one, $\mathbf{I}_N$ is an $N \times N$ identity matrix and $\otimes$ denotes the Kronecker product. We assume that a two-stage simple random sample is to be selected (without replacement) from this population. At the first stage, a sample of n clusters is selected and at a second stage, a sample of m distinct units is selected from the M units in each selected cluster.

2.1. The finite population mixed (FM) model

We briefly summarize the *FM* model presented by Stanek and Singer (2004). This model is a stochastic version of (2.1) induced by the two-stage random sampling of a finite clustered population. To obtain the *FM* model, we first define an ordered list of *NM* random variables, the values of which correspond to the elements of $\mathbf{y}$ obtained from independent permutations of clusters and units in clusters. For each permutation, we assign a new label, $i = 1, \ldots, N$ to the clusters according to its position in the permuted list. Similarly, we label the positions in the permutation of units in a cluster by $j = 1, \ldots, M$. For ease of exposition, we refer to the cluster that will occupy position i in the permutation of clusters as primary sampling unit (*PSU*) i, and to the unit that will occupy position j in the permutation of units within a cluster as secondary sampling unit (*SSU*) j. Since any unit in any cluster may occupy position ij, we represent the response for *SSU j* in *PSU i* as the random variable Y_{ij}.

To relate y_{st} to Y_{ij} we use two sets of indicator random variables; the elements of the first, U_{is}, take on a value of one when the realized cluster corresponding to *PSU i* is cluster*s* and a value of zero otherwise, and the elements of the second set, $U_{jt}^{(s)}$, take on a value of one when the realized unit corresponding to *SSU j* in cluster s is unit t and zero otherwise. As a consequence, the random variable corresponding to the response for *SSU j* in *PSU i* in a permutation is given by

$$Y_{ij} = \sum_{s=1}^{N}\sum_{t=1}^{M} U_{is}U_{jt}^{(s)}y_{st}.$$

The finite population $\mathbf{y}$ can be viewed as a realization of the random variable

$$\mathbf{Y} = (\mathbf{U} \otimes \mathbf{I}_M) \left(\bigoplus_{s=1}^{N} \mathbf{U}^{(s)} \right) \mathbf{y},$$

where $\mathbf{Y}=(\mathbf{Y}_1' \quad \mathbf{Y}_2' \quad \cdots \quad \mathbf{Y}_N')' \in \mathbb{R}^{NM}$, with $\mathbf{Y}_i=(Y_{i1} \quad Y_{i2} \quad \cdots \quad Y_{iM})' \in \mathbb{R}^M$, $\mathbf{U}^{(s)}=(\mathbf{U}_1^{(s)} \quad \mathbf{U}_2^{(s)} \quad \cdots \quad \mathbf{U}_M^{(s)}) \in \mathbb{R}^M \times \mathbb{R}^M$, with $\mathbf{U}_t^{(s)} = (U_{1t}^{(s)} \; U_{2t}^{(s)} \; \cdots \; U_{Mt}^{(s)})'$, and $\mathbf{U} = (\mathbf{U}_1 \quad \mathbf{U}_2 \quad \cdots \quad \mathbf{U}_N) \in \mathbb{R}^N \times \mathbb{R}^N$, with columns $\mathbf{U}_s = (U_{1s} \quad U_{2s} \quad \cdots \quad U_{Ns})'$. Here, $\oplus_{s=1}^N \mathbf{A}_s$ denotes a block diagonal matrix with blocks $\mathbf{A}_s$ (Searle et al., 1992). One difference between the vectors $\mathbf{Y}$and $\mathbf{y}$ lies in the interpretation of the subscripts that define their elements. The subscripts in $\mathbf{Y}$ correspond to positions in a permutation, while the subscripts in $\mathbf{y}$ correspond to labels of units.

Using elementary properties of the indicator random variables and the simple structure of $\mathbf{X}$ and $\mathbf{Z}$, it follows that $\mathbf{U}\mathbf{1}_N=\mathbf{1}_N$, $\mathbf{U}^{(s)}\mathbf{1}_M=\mathbf{1}_M$, $(\mathbf{U}\otimes\mathbf{I}_M)(\oplus_{s=1}^N \mathbf{U}^{(s)})\mathbf{X}=\mathbf{X}$ and $(\mathbf{U}\otimes\mathbf{I}_M)(\oplus_{s=1}^N \mathbf{U}^{(s)})\mathbf{Z}=\mathbf{U}\otimes\mathbf{1}_M=\mathbf{Z}\mathbf{U}$. Then, pre-multiplying both sides of (2.1) by $(\mathbf{U} \otimes \mathbf{I}_M) \left(\bigoplus_{s=1}^{N} \mathbf{U}^{(s)} \right)$ and using the above results, we obtain the *FM* model

$$\mathbf{Y} = \mathbf{X}\mu + \mathbf{Z}\mathbf{B} + \mathbf{E}, \tag{2.2}$$

where $\mathbf{E} = (\mathbf{U} \otimes \mathbf{I}_M)(\oplus_{s=1}^N \mathbf{U}^{(s)})\boldsymbol{\varepsilon}$ and $\mathbf{B} = \mathbf{U}\boldsymbol{\beta} = (B_1 \quad B_2 \quad \cdots \quad B_N)'$. Given the random nature of $\mathbf{U}$, the term $B_i = \sum_{s=1}^N U_{is}\beta_s$ is a random effect that represents the deviation of the latent value for *PSU* i from the population mean.

The expected value and covariance matrix of $\mathbf{Y}$ can be developed directly from the properties of the sampling indicator random variables which formally represent the two-stage cluster sampling, as illustrated by Stanek and Singer (2004, Appendix A). For the random variable $\mathbf{Y}$ in (2.2), we have

$$E_{\xi_1\xi_2}(\mathbf{Y}) = \mathbf{X}\mu$$

and

$$\mathrm{var}_{\xi_1\xi_2}(\mathbf{Y}) = \sigma_e^2\mathbf{I}_{NM} + \sigma^{*2}(\mathbf{I}_N \otimes \mathbf{J}_M) - \frac{\sigma^2}{N}\mathbf{J}_{NM},$$

where $\sigma^{*2} = \sigma^2 - \sigma_e^2/M$, $\mathbf{J}_a = \mathbf{1}_a\mathbf{1}_a'$, and the subscripts ξ_1 and ξ_2 denote expectation with respect to permutations of the clusters and to permutations of units in the clusters, respectively.

As the values of the finite population defined by $\mathbf{y}$ may not be observed directly, we assume a model of the form

$$\tilde{\mathbf{Y}} = \mathbf{y} + \mathbf{W},$$

where $\mathbf{W} = (W_{st}) \in \mathbb{R}^{NM}$ is a vector of independent response errors with $E(W_{st}) = 0$ and $\mathrm{var}(W_{st}) = \sigma_{st}^2$, $s = 1, \ldots,$ $t = 1, \ldots, M$.

Using this notation, the *FM* model with response error is

$$\mathbf{Y}^* = (\mathbf{U} \otimes \mathbf{I}_M) \left(\bigoplus_{s=1}^{N} \mathbf{U}^{(s)} \right) \tilde{\mathbf{Y}} = \mathbf{Y} + \mathbf{W}^*,$$

where $\mathbf{W}^* = (\mathbf{U} \otimes \mathbf{I}_M)(\oplus_{s=1}^N \mathbf{U}^{(s)})\mathbf{W}$. Under the re-parameterization (2.1), we express this as

$$\mathbf{Y}^* = \mathbf{X}\mu + \mathbf{Z}\mathbf{B} + (\mathbf{E} + \mathbf{W}^*). \tag{2.3}$$

The first two central moments of $\mathbf{Y}^*$ are

$$E_{\xi_1\xi_2\xi_3}(\mathbf{Y}^*) = \mathbf{X}\mu$$

and

$$\mathrm{var}_{\xi_1\xi_2\xi_3}(\mathbf{Y}^*) = (\sigma_e^2 + \sigma_r^2)\mathbf{I}_{NM} + \sigma^{*2}(\mathbf{I}_N \otimes \mathbf{J}_M) - \frac{\sigma^2}{N}\mathbf{J}_{NM}.$$

Here, the subscript ξ_3 denotes expectation with respect to response error and $\sigma_r^2 = \sum_{s=1}^N \sum_{t=1}^M \sigma_{st}^2/NM$ denotes the average response error variance.

The same finite population with two-stage permuted elements is represented in each realization of $\mathbf{Y}^*$. Once the sample has been selected, the correspondence between clusters and particular sample *PSU*s will be apparent. We refer to those clusters as realized *PSU*s.

Denoting the sample elements by $\mathbf{Y}_I^* = \mathbf{Y}_I + \mathbf{W}_I^*$, the model for the sample is $\mathbf{Y}_I^* = \mathbf{X}_I\mu + \mathbf{Z}_I\mathbf{B} + (\mathbf{E}_I + \mathbf{W}_I^*)$ where $\mathbf{X}_I = \mathbf{1}_n \otimes \mathbf{1}_m$, $\mathbf{Z}_I = \mathbf{I}_n \otimes \mathbf{1}_m$, $E(\mathbf{Y}_t^*) = \mathbf{X}_I\mu$ and $\mathrm{var}(\mathbf{Y}_I^*) = (\sigma_e^2 + \sigma_r^2)\mathbf{I}_{nm} + \sigma^{*2}(\mathbf{I}_n \otimes \mathbf{J}_m) - (\sigma^2/N)\mathbf{J}_{nm}$.

2.2. Scott and Smith's SP model

Scott and Smith (1969) proposed an *SP* model according to which the finite population $\mathbf{y}$ is viewed as a realization of a vector of random variables $\mathbf{Y}$ such that

$$E(\mathbf{Y}) = \mathbf{X}\mu \quad \text{and} \quad \mathrm{var}(\mathbf{Y}) = \bigoplus_{i=1}^{N}(\sigma_i^2\mathbf{I}_M + \sigma^2\mathbf{J}_M). \tag{2.4}$$

Although we use the same notation, neither μ nor σ^2 and σ_i^2 necessarily refer to the finite population mean or variance components specified at the beginning of Section 2, because the vector $\mathbf{Y}$ in (2.4) is not linked to the population units as in the *FM* model. In this context, the cluster means, i.e., $\mu_1, \ldots, \mu_N$ may be considered as realizations of independent identically distributed random variables $\Lambda_1, \ldots, \Lambda_N$ in a SP such that for each $i = 1, \ldots, N$, $E(\Lambda_i) = \mu$ and $\mathrm{var}(\Lambda_i) = \sigma^2$. According to this model, elements within the same cluster are correlated, but elements in different clusters are not.

Using a Bayesian approach, Bolfarine and Zacks (1992) extend the *SP* model by adding normally distributed response error, and suppose that the potentially observable variables are given by elements of

$$\mathbf{Y}^* = \mathbf{Y} + \mathbf{W},$$

where $\mathbf{W} \sim N(\mathbf{0}, \oplus_{i=1}^{N}\sigma_{ri}^2\mathbf{I}_M)$, $i = 1, \ldots, N$, $j = 1, \ldots, M$, and $\mathbf{W}$ is independent of $\mathbf{Y}$. Assuming that $\sigma_{ri}^2 = \sigma_r^2$, $i = 1, \ldots, N$, we obtain

$$E(\mathbf{Y}^*) = \mathbf{X}\mu \quad \text{and} \quad \mathrm{var}(\mathbf{Y}^*) = \bigoplus_{i=1}^{N}[(\sigma_i^2 + \sigma_r^2)\mathbf{I}_M + \sigma^2\mathbf{J}_M]. \tag{2.5}$$

Denoting the sample elements by $\mathbf{Y}_t^* = \mathbf{Y}_I + \mathbf{W}_I$, it follows that $E(\mathbf{Y}_I^*) = \mathbf{X}_I\mu$ and

$$\mathrm{var}(\mathbf{Y}_I^*) = \bigoplus_{i=1}^{n}[(\sigma_i^2 + \sigma_r^2)\mathbf{I}_m + \sigma^2\mathbf{J}_m].$$

2.3. The mixed effects (ME) model

Under an *ME* model, the two-stage sample data are considered to have been selected from a conceptual infinite population, which may be thought of as the limit (as the size increases) of the finite population of interest. In this case, the sample elements in *PSU i*, given by $\mathbf{Y}_{Ii} = (Y_{i1}\ Y_{i2}\ \cdots\ Y_{im})'$, $i = 1, \ldots, n$, are modeled as

$$\mathbf{Y}_{Ii} = \mathbf{X}_{Ii}\mu + \mathbf{Z}_{Ii}B_i + \mathbf{E}_{Ii}, \tag{2.6}$$

where $\mathbf{X}_{Ii} = \mathbf{Z}_{Ii} = \mathbf{1}_m$ and $\mathbf{E}_{Ii} = (E_{i1}\ E_{i2}\ \cdots\ E_{im})'$. Here, Y_{ij} is the response of *SSU j*, $j = 1, \ldots, m$ in *PSU i*, μ corresponds to the expected response over *SSU*s and *PSU*s in the conceptual infinite population, B_i is a random effect that corresponds to the deviation of the average expected response of *SSU*s in *PSU i* from μ, and E_{ij} is a random deviation of the (conditional) expected response of *SSU j* from the (conditional) average expected response of elements in *PSU i*. Typically, it is assumed that $E(B_i) = 0$, $\mathrm{var}(B_i) = \sigma^2$, $E(\mathbf{E}_{Ii}) = \mathbf{0}$, $\mathrm{var}(\mathbf{E}_{Ii}) = \sigma_i^2\mathbf{I}_m$ and $\mathrm{cov}(B_i, E_{ij}) = 0$, $j = 1, \ldots, m$, so that $E(\mathbf{Y}_{Ii}) = \mathbf{X}_{Ii}\mu$ and $\mathrm{var}(\mathbf{Y}_{Ii}) = \sigma_i^2\mathbf{I}_m + \sigma^2\mathbf{J}_m$. If to (2.6) we add a response error term $\mathbf{W}_{Ii} = (W_{i1}\ W_{i2}\ \cdots W_{im})'$, with $E(\mathbf{W}_{Ii}) = \mathbf{0}$, $\mathrm{var}(\mathbf{W}_{Ii}) = \sigma_r^2\mathbf{I}_m$, $\mathrm{cov}(B_i, W_{ij}) = 0$, $j = 1, \ldots, m$ and $\mathrm{cov}(\mathbf{W}_{Ii}, \mathbf{E}_{Ii}) = \mathbf{0}$, then for $\mathbf{Y}_{Ii}^* = \mathbf{Y}_{Ii} + \mathbf{W}_{Ii}$, we have $E(\mathbf{Y}_{Ii}^*) = \mathbf{X}_{Ii}\mu$ and $\mathrm{var}(\mathbf{Y}_{Ii}^*) = (\sigma_i^2 + \sigma_r^2)\mathbf{I}_m + \sigma^2\mathbf{J}_m$. It follows that

$$\mathbf{Y}_I^* = (\mathbf{Y}_{I1}^{*'}\ \mathbf{Y}_{I2}^{*'}\ \cdots\ \mathbf{Y}_{In}^{*'})' = \mathbf{X}_I\mu + \mathbf{Z}_I\mathbf{B}_I + \mathbf{E}_I + \mathbf{W}_I, \tag{2.7}$$

where $\mathbf{B}_I = (\,B_1 \quad B_2 \quad \cdots \quad B_n\,)'$, $\mathbf{E}_I = (\mathbf{E}'_{I1} \quad \mathbf{E}'_{I2} \quad \cdots \quad \mathbf{E}'_{In}\,)'$, and $\mathbf{W}_I = (\mathbf{W}'_{I1} \quad \mathbf{W}'_{I2} \quad \cdots \quad \mathbf{W}'_{In}\,)'$, which implies that $E(\mathbf{Y}^*_I) = \mathbf{X}_I\mu$ and $\text{var}(\mathbf{Y}^*_I) = \oplus_{I=1}^{n}\,[(\sigma_i^2 + \sigma_r^2)\mathbf{I}_m + \sigma^2\mathbf{J}_m]$.

Here again, neither μ nor σ^2 and σ_i^2 refer directly to the finite population mean or variance components, but to the conceptual infinite population counterparts. In particular, σ^2 is the variance of the random cluster means $(\mu + B_i)$ which conceptually take on an infinite number of values as opposed to the finite population setup, where they can assume only a finite number of values.

2.4. BLUPs of the latent value of a realized PSU

Our principal interest lies in the linear combination that defines the latent value of *PSU i* (for $i \leqslant n$), i.e., $T_i = \mathbf{g}'\mathbf{Y}$, where $\mathbf{g}' = (1/M)\mathbf{e}'_i \otimes \mathbf{1}'_M$, and $\mathbf{e}_i$ denotes an $N \times 1$ column vector with 1 in position i and zero elsewhere. In the *ME* model, T_i corresponds to $\mu + B_i$. From the sampled values $\mathbf{Y}^*_I = (\mathbf{Y}^{*'}_{I1}\ \mathbf{Y}^{*'}_{I2}\ \cdots\ \mathbf{Y}^{*'}_{In})'$, where $\mathbf{Y}^*_{Ii} = (\,Y^*_{i1} \quad Y^*_{i2} \quad \cdots \quad Y^*_{im}\,)'$, the BLUPs of T_i (for $i \leqslant n$) under the three models [i.e., (2.3), (2.5) or (2.7)] are:

(i) for the *ME* model:

$$\hat{T}_i = \hat{\mu} + k_i^{(ME)}(\bar{Y}^*_i - \hat{\mu}) \quad \text{where } \bar{Y}^*_i = \frac{1}{m}\sum_{j=1}^{m} Y^*_{ij},\ \ \hat{\mu} = \frac{\sum_{i=1}^{n}\bar{Y}^*_i/v_i}{\sum_{i=1}^{n}1/v_i},$$

$v_i = m\sigma^2 + \sigma_i^2 + \sigma_r^2$, and the shrinkage constant is

$$k_i^{(ME)} = \frac{m\sigma^2}{m\sigma^2 + \sigma_i^2 + \sigma_r^2}.$$

(ii) for the *SP* model: $\hat{T}_i = \hat{\mu} + k_i^{(SP)}(\bar{Y}^*_i - \hat{\mu})$, and the shrinkage constant is

$$k_i^{(SP)} = k_i^{(ME)} + f(1 - k_i^{(ME)}) = \frac{m\sigma^2 + f(\sigma_i^2 + \sigma_r^2)}{m\sigma^2 + \sigma_i^2 + \sigma_r^2},$$

with $f = m/M$ denoting the sampling fraction for units.

(iii) for the *FM* model: $\hat{T}_i = \bar{Y}^* + k^{(FM)}(\bar{Y}^*_i - \bar{Y}^*)$, where $\bar{Y}^* = (1/n)\sum_{i=1}^{n}\bar{Y}^*_i$ and the shrinkage constant is

$$k^{(FM)} = \frac{m\sigma^2}{m\sigma^{*2} + \sigma_e^2 + \sigma_r^2} = \frac{m\sigma^2}{m\sigma^2 + (1-f)\sigma_e^2 + \sigma_r^2}.$$

Under the assumption that the within-cluster variances are identical for all clusters (and equal to σ_e^2), the BLUPs of T_i (for $i \leqslant n$) under the three models reduce to

$$\hat{T}_i = \bar{Y}^* + k^{(\text{model})}(\bar{Y}^*_i - \bar{Y}^*),$$

where

(i) for the *ME* model:

$$k^{(ME)} = \frac{m\sigma^2}{m\sigma^2 + \sigma_e^2 + \sigma_r^2},$$

(ii) for the *SP* model:

$$k^{(SP)} = \frac{m\sigma^2 + f(\sigma_e^2 + \sigma_r^2)}{m\sigma^2 + \sigma_e^2 + \sigma_r^2},$$

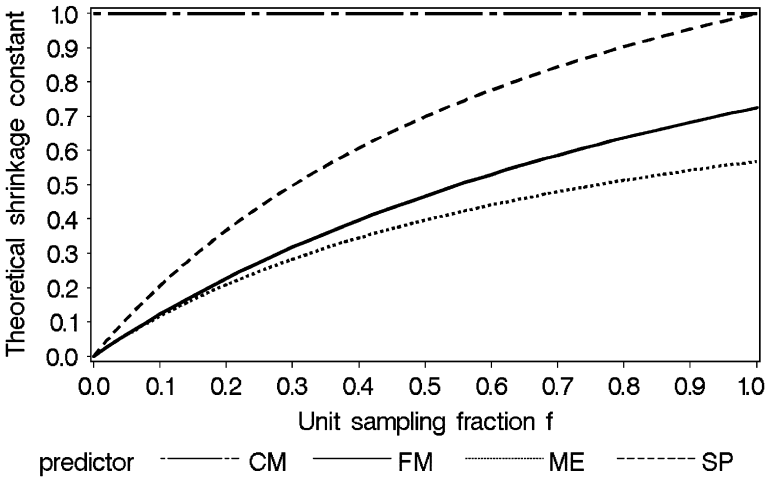


Fig. 1. Theoretical shrinkage constants for different values of unit sampling fractions, with cluster intra-class correlation coefficient $\rho_s = 0.05$ and unit intra-class correlation coefficient $\rho_t = 0.5$.

(iii) for the *FM* model:

$$k^{(FM)} = \frac{m\sigma^2}{m\sigma^{*2} + \sigma_e^2 + \sigma_r^2} = \frac{m\sigma^2}{m\sigma^2 + (1-f)\sigma_e^2 + \sigma_r^2}.$$

It follows that $0 \leqslant k^{(ME)} \leqslant k^{(FM)} \leqslant k^{(SP)} \leqslant 1$.

Studying the behavior of these theoretical shrinkage constants is a first step to understand the similarities and differences between the predictors, although they do not take into account the cluster sampling fraction, as the *MSE* does. In Fig. 1, we compare the behavior of the theoretical shrinkage constants as f varies, for a cluster intra-class correlation $\rho_s = 0.05$ and a unit intra-class correlation[2] $\rho_t = 0.5$. We include the performance of the cluster mean (*CM*) for which the shrinkage constant $k_2^{(CM)}$ is always equal to 1 (dot-dashed line). The solid line identifies the theoretical shrinkage constant for the *FM* model predictor while the dashed and the dotted lines identify the theoretical shrinkage constants for the *SP* and *ME* model predictors, respectively. From Fig. 1 we observe that when f tends to 0, all the shrinkage constants (except $k^{(CM)}$) become more similar and also that they increase as the unit sampling fraction increases.

In Fig. 2, we compare the behavior of the theoretical shrinkage constants as f, ρ_s and ρ_t vary. Labels are similar to those of Fig. 1 and are omitted for visual clarity. Since the *FM* model predictors have smallest *MSE* when variance components are known, we consider it as reference to compare the rivals and identify those with similar as well as different (poor) performance relatively to it.

We expect minor differences between the predictors, almost independently of f, when both ρ_s and ρ_t tend to one. In Fig. 2, we also observe that when f tends to 0, $k^{(ME)}$, $k^{(SP)}$ and $k^{(FM)}$ become more similar (and all of them differ from $k^{(CM)}$). When there is no response error ($\rho_t = 1$) and all units in a cluster are sampled ($f = 1$), $k^{(FM)} = k^{(SP)} = k^{(CM)} = 1$ (and all differ from $k^{(ME)}$), so that the best predictor is the *CM*. When ρ_t tends to zero, $k^{(ME)}$ approaches $k^{(FM)}$ (almost independently of f), so that for this situation, we expect the predictors derived under these two models to behave similarly.

Relatively to the theoretical results for the *FM* model predictor, when both ρ_s and ρ_t tend to zero, we expect the *CM* to have poor performance when f approaches zero, while the *SP* model predictor is expected to have poor performance as f approaches one. Also, we expect the *ME* model predictor to have poor performance as ρ_s tends to zero and both ρ_t and f tend to one.

[2] We define the cluster intra-class correlation coefficient as $\rho_s = \sigma^2/(\sigma^2 + \sigma_e^2)$ and the unit intra-class correlation coefficient as $\rho_t = \sigma_e^2/(\sigma_e^2 + \sigma_r^2)$.

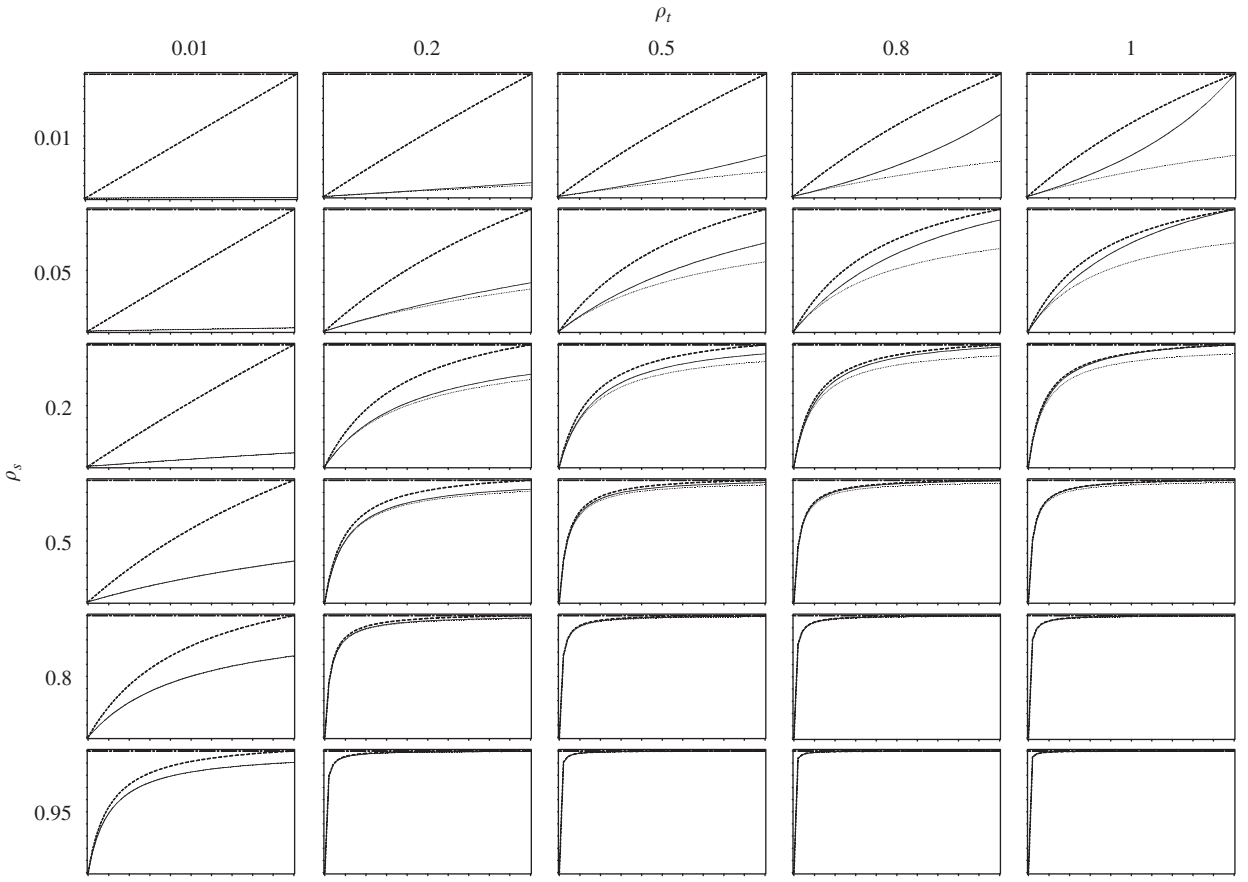


Fig. 2. Theoretical shrinkage constants for different values of intra-class correlation coefficients and unit sampling fractions.

2.5. Empirical BLUPs of the latent value of a realized PSU

In practice, variance components are usually unknown and estimates are needed for the shrinkage constants. Empirical predictors can be obtained by substituting the shrinkage constants by their respective estimators. To estimate variance components, Searle and Fawcett (1970) developed a rule for converting expectations of mean squares obtained under variance components infinite population models into expectations under finite population models, but these rules have been seldom used, due, in part, to a lack of additional theoretical results and software. As the finite population models used in this work do not involve any assumption about the response distribution besides the existence and the structure of the first two central moments, we use method of moments to obtain estimators for the variance components. These estimators may be derived from ANOVA mean squares, namely $MSB = (n-1)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}(\bar{Y}_i^* - \bar{Y}^*)^2$ and $MSR = [n(m-1)]^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij}^* - \bar{Y}_i^*)^2$ that may be expressed as quadratic forms of the type $\mathbf{Y}_I^{*'}\mathbf{A}\mathbf{Y}_I^*$ where $\mathbf{A}$ is a convenient matrix. In particular, $\mathbf{A} = (n-1)^{-1}(\mathbf{P}_n \otimes \mathbf{J}_m/m)$ for *MSB* and $\mathbf{A} = [n(m-1)]^{-1}(\mathbf{I}_n \otimes \mathbf{P}_m)$ for *MSR*, with $\mathbf{P}_a = \mathbf{I}_a - a^{-1}\mathbf{J}_a$ and a denotes a positive integer. Since $\mathbf{A}\mathbf{1}_{nm} = \mathbf{0}$ and $E(\mathbf{Y}_I^*) = \mathbf{1}_{nm}\mu$, under any of the three models, it follows that $E(\mathbf{Y}_I^{*'})\mathbf{A}E(\mathbf{Y}_I^*) = 0$ and therefore,

$$E(\mathbf{Y}_I^{*'}\mathbf{A}\mathbf{Y}_I^*) = \mathrm{tr}[\mathrm{var}(\mathbf{Y}_I^*)\mathbf{A}] + E(\mathbf{Y}_I^{*'})\mathbf{A}E(\mathbf{Y}_I^*) = \mathrm{tr}[\mathrm{var}(\mathbf{Y}_I^*)\mathbf{A}]. \tag{2.8}$$

Using these results, we obtain estimators for the variance components under each of the three competing models, derive estimators for the shrinkage constants and obtain the corresponding empirical predictors as described next.

2.5.1. Empirical BLUP under the FM model

Under the *FM* model, we have

$$\mathrm{var}(\mathbf{Y}_I^*) = (\sigma_e^2 + \sigma_r^2)\mathbf{I}_{nm} + \sigma^{*2}(\mathbf{I}_n \otimes \mathbf{J}_m) - \frac{\sigma^2}{N}\mathbf{J}_{nm}.$$

To evaluate the expected value of *MSR*, we let $\mathbf{A}=[n(m-1)]^{-1}(\mathbf{I}_n \otimes \mathbf{P}_m)$ in (2.8) which implies that $E(MSR)=\sigma_e^2+\sigma_r^2$. To evaluate the expected value of *MSB*, we let $\mathbf{A} = (n-1)^{-1}(\mathbf{P}_n \otimes \mathbf{J}_m/m)$ in (2.8) which implies that

$$E(MSB) = m\sigma^{*2} + \sigma_e^2 + \sigma_r^2 = m\sigma^2 + (1-f)\sigma_e^2 + \sigma_r^2.$$

Assuming that the response error variance σ_r^2 is known and equating the observed and expected mean squares, we obtain $\hat{\sigma}_e^2 + \sigma_r^2 = MSR$ and $m\hat{\sigma}^2 = MSB - (1-f)MSR - f\sigma_r^2$. Consequently, one method of moments estimator for $k^{(FM)}$ is

$$\hat{k}_1^{(FM)} = \begin{cases} 0 & \text{if } MSB = 0, \\ \max\left(0, \dfrac{MSB-(1-f)MSR-f\sigma_r^2}{MSB}\right) & \text{if } MSB > 0. \end{cases}$$

Re-expressing $k^{(FM)}$ as

$$k^{(FM)} = \frac{m\sigma^2}{m\sigma^{*2} + \sigma_e^2 + \sigma_r^2} = \frac{m\sigma^{*2} + f\sigma_e^2}{m\sigma^{*2} + \sigma_e^2 + \sigma_r^2} = \frac{m\sigma^{*2} + f\rho_t(\sigma_e^2 + \sigma_r^2)}{m\sigma^{*2} + \sigma_e^2 + \sigma_r^2},$$

an alternative method of moments estimator may be obtained when ρ_t is assumed known. Equating the observed and expected mean squares, it follows that $m\hat{\sigma}^{*2} + \hat{\sigma}_e^2 + \hat{\sigma}_r^2 = MSB$, $\hat{\sigma}_e^2 + \hat{\sigma}_r^2 = MSR$ and $m\hat{\sigma}^{*2} = MSB–MSR$. Then, an alternative method of moments estimator for $k^{(FM)}$ is

$$\hat{k}_2^{(FM)} = \begin{cases} 0 & \text{if } MSB = 0, \\ \max\left(0, \dfrac{MSB-(1-f\rho_t)MSR}{MSB}\right) & \text{if } MSB > 0. \end{cases}$$

2.5.2. Empirical BLUPs under the ME and SP models

Under the *ME* and *SP* models, we have

$$\mathrm{var}(\mathbf{Y}_I^*) = \bigoplus_{i=1}^{n}(a_i^2\mathbf{I}_m + b^2\mathbf{J}_m),$$

where $a_i^2 = \sigma_i^2 + \sigma_r^2$, and $b^2 = \sigma^2$. Using $\mathbf{A} = [n(m-1)]^{-1}(\mathbf{I}_n \otimes \mathbf{P}_m)$ in (2.8) it follows that

$$E(MSR) = \frac{1}{n}\sum_{i=1}^{n} a_i^2.$$

Letting $\mathbf{A} = (n-1)^{-1}(\mathbf{P}_n \otimes \mathbf{J}_m/m)$ in (2.8) it follows that

$$E(MSB) = \frac{1}{n}\sum_{i=1}^{n} a_i^2 + mb^2.$$

When within-cluster variances are equal (i.e., $\sigma_i^2 = \sigma_e^2$ for all $i = 1, \ldots, n$), we have $a_i^2 = \sigma_e^2 + \sigma_r^2$ for all $i = 1, \ldots, n$, and the expected mean squares terms reduce to

$$E(MSR) = \sigma_e^2 + \sigma_r^2$$

and

$$E(MSB) = \sigma_e^2 + \sigma_r^2 + m\sigma^2.$$

In this context, equating observed and expected mean squares, we obtain

$$\hat{k}^{(ME)} = \frac{m\hat{\sigma}^2}{m\hat{\sigma}^2 + \hat{\sigma}_e^2 + \hat{\sigma}_r^2} = \begin{cases} 0 & \text{if } MSB = 0, \\ \max\left(0, \frac{MSB - MSR}{MSB}\right) & \text{if } MSB > 0 \end{cases}$$

for the *ME* model,[3] and

$$\hat{k}_1^{(SP)} = \hat{k}^{(ME)} + f(1 - \hat{k}^{(ME)}) = \begin{cases} f & \text{if } 0 \leqslant MSB \leqslant MSR, \\ \frac{MSB - (1-f)MSR}{MSB} & \text{if } MSB > MSR \end{cases}$$

or

$$\hat{k}_2^{(SP)} = \frac{m\hat{\sigma}^2 + f(\hat{\sigma}_e^2 + \hat{\sigma}_r^2)}{m\hat{\sigma}^2 + \hat{\sigma}_e^2 + \hat{\sigma}_r^2} = \begin{cases} f & \text{if } MSB = 0, \\ \max\left(0, \frac{MSB - (1-f)MSR}{MSB}\right) & \text{if } MSB > 0 \end{cases}$$

for the *SP* model.

The empirical shrinkage constants satisfy $0 \leqslant \hat{k}^{(ME)} \leqslant \hat{k}_2^{(FM)} \leqslant \hat{k}_2^{(SP)} \leqslant 1$, $0 \leqslant \hat{k}_1^{(FM)} \leqslant \hat{k}_2^{(SP)} \leqslant 1$ and $0 \leqslant \hat{k}^{(ME)} \leqslant \hat{k}_1^{(SP)} \leqslant 1$.

3. Details of the simulation study

We conducted a simulation study to compare the *MSE* of the different empirical predictors in the context of a two-stage cluster sample from a balanced finite population. The simulation study was carried out in three steps: (1) generation of the finite population, (2) selection of two-stage cluster samples from the finite population, and (3) evaluation of predictors and empirical predictors of the realized *PSU* latent values for comparative purposes. The first two steps provide a common context where the three models have been applied. We evaluated the predictors in this context, without regard to the prior distribution from which the finite population parameters were obtained. The results reflect what would be expected from repeated two-stage sampling of a finite population, and do not depend on the prior distribution or other parametric assumptions. Superpopulation model approaches have been used to evaluate competing predictors based on simulated clustered population settings (Li and Lahiri, 2006, Pfeffermann and Nathan, 1981), but the superpopulation models used have not simultaneously included lack of fit and response error. We base the simulation studies on the actual physical process that underlies the sampling design and measurement. This approach accounts simultaneously for lack of fit, response error and the two-stage sampling without additional assumptions.

3.1. Generation of the finite populations

To encompass a broad number of situations, different compositions for the finite populations were considered. They differ with respect to: (1) the number of clusters, N and the number of units within clusters, M, (2) the shape of the response distribution, and (3) the between-cluster variance, σ^2. The presence (or absence) of response error was considered at the sampling stage.

Each population of units and clusters was generated via the percentiles of some hypothetical distribution. The basic distributions used for the finite populations were the normal, uniform, beta, and gamma. These distributions were used only to generate the cluster means; their actual form was not used in the analysis. Although different distributions can

[3] Note that if we assume that the response error variance (σ_r^2) is known, we would obtain the same estimator for the shrinkage constant, which makes no use of this information.

Table 1
Characteristics of the simulated populations and sampling plans

Case	No. of clusters N	No. of units per cluster M	Cluster parameter distribution	Unit parameter distribution	No. of cluster intra-class correlation coefficients: ρ_s [1]	No. of Unit intra-class correlation coefficients: ρ_t [2]	No. of generated populations	No. of cluster sampling fractions F [3]	No. of unit sampling fractions f [4]	No. Settings
1	10	5	Normal	Normal	7	8	56	3	3	504
2	10	5	Normal	Normal Uniform Beta(10,1) Beta(0.5,0.5) Gamma(0.5) Gamma(2)	7	8	336	3	3	3024
3	10	5	Normal Uniform Beta(10,1) Beta(0.5,0.5) Gamma(0.5) Gamma(2)	Normal	7	8	336	3	3	3024
4	10	20	Normal	Normal	7	8	56	3	7	1176
5	50	20	Normal	Normal	7	8	56	3	7	1176

[1] $\rho_s = 0.01, 0.05, 0.2, 0.5, 0.8, 0.95, 0.99$.
[2] $\rho_t = 0.01, 0.05, 0.2, 0.5, 0.8, 0.95, 0.99, 1$.
[3] $F = 0.2, 0.5, 0.8$.
[4] $f = 0.4, 0.6, 0.8$ or $0.1, 0.2, 0.4, 0.5, 0.6, 0.8, 0.9$.

be selected for units within clusters, we used the same distribution to generate the unit effects for all clusters in each population. The cluster distribution may or may not agree with the distribution adopted for the units.

For each simulation, the population was composed of N clusters with M units per cluster. We represented each individual cluster parameter by μ_s and their mean by μ. We fixed the between-cluster parameters variance, σ^2, divided the [0,1] interval into $N + 1$ equally spaced subintervals and obtained the percentiles corresponding to the upper limit of each subinterval from the appropriate probability distribution. We redefined the cluster parameters by centering them at μ and re-scaling their values so that the variance matched

$$\sigma^2 = \sum_{s=1}^{N} \frac{(\mu_s - \mu)^2}{N - 1}.$$

Next, we generated unit effects for the M units within each cluster, using percentiles of a specified distribution. The variance of the unit effects was set constant for all clusters. Unit effects were then re-scaled to have zero mean for each cluster and within-cluster variance equal to a specified value, σ_e^2. The unit parameters, y_{st}, were formed by adding the unit effects to the cluster mean.

Using σ^2, σ_e^2 and σ_r^2, we defined the cluster and unit intra-class correlation coefficients. We assumed that the response error was normally distributed for all clusters and units, with the response variance determined by specification of ρ_t and σ_e^2. Note that $\rho_t = 1$ corresponds to the case with no response error.

The characteristics of the simulated populations are summarized in Table 1. Fifty-six populations were generated in case 1, corresponding to combinations of seven cluster and eight unit intra-class correlation coefficients. In cases 2 and 3, $336 = (6 \times 7 \times 8)$ populations were generated and in cases 4 and 5, $56 = (7 \times 8)$ populations were generated. In total, 840 populations were evaluated.

3.2. Two-stage cluster sampling

A simple random sample without replacement of n cluster labels obtained from a list was identified for each generated population. The identified sample cluster labels were combined with the population data, and following a similar process, a simple random sample without replacement of m units in each sampled cluster was selected from the labeled data. When response error was considered, it was added to the unit parameters y_{st} during the selection of the second-stage samples. We refer to this entire process as a 'trial'.

For each population generated under cases 1, 2, and 3 (Table 1), three cluster sampling fractions ($F = n/N = 0.2$, 0.5 and 0.8) and three unit sampling fractions ($f = m/M = 0.4$, 0.6 and 0.8) were considered. This resulted in nine sampling plans for each generated population. For cases 4 and 5 (Table 1), three cluster sampling fractions ($F = 0.2$, 0.5 and 0.8) and seven unit sampling fractions ($f = 0.1$, 0.2, 0.4, 0.5, 0.6, 0.8 and 0.9) were considered, resulting in 21 sampling plans for each generated population. For each population and sampling plan, the number of trials was 10 000. In total, 8904 different settings for population/sampling plans were simulated.

3.3. BLUPs and empirical BLUPs of the realized PSU latent values

Once the two-stage samples were obtained, we computed the predictors and empirical predictors of the realized *PSU* latent values as detailed in Sections 2.4 and 2.5, respectively. When variances were considered known, the *ME* and *SP* model predictors were obtained under the assumption that the values of σ^2 and σ_e^2 corresponded to the between and within-cluster variances, respectively. The *MSE* of each predictor was obtained as the mean of the squared differences between the predictor and the true value of the realized *PSU* latent value. For the cases where variance components were considered known, the observed *MSE* was denoted by *SMSE*; otherwise, the observed *MSE* was denoted by *EMSE*. The latter correspond to the empirical predictors described in Section 2.5

4. Simulation results

To clarify the exposition, we present the results in two sections. First, we evaluate the performance of each predictor (i.e., with known variance components) by comparing their *SMSE* in order to give a reference framework for the empirical predictor results. Second, we evaluate the performance of the empirical predictors (i.e., with estimated variance components).

Initially, we compute the relative loss in terms of *EMSE* with respect to *SMSE* that occurs when we replace the theoretical shrinkage constants by their estimators obtained under each of the three competing models (i.e., when using the empirical predictors). Then, we determine under what settings each empirical predictor presents the best performance as well as under what settings they perform poorly.

To compare the (empirical) predictors we consider three criteria. First, we identify the best (empirical) predictor for each setting as the one that presents minimum (*EMSE*) *SMSE*. As sometimes the differences between the (*EMSE*) *SMSE* of two (empirical) predictors is very small, we also use the relative percent increase[4] (*RPI*) in (*EMSE*) *SMSE* of each (empirical) predictor relatively to the (*EMSE*) *SMSE* of the best (empirical) predictor to identify settings where two (empirical) predictors may be considered "equivalent". We use $RPI < 5\%$ or $RPI < 15\%$ as a criterion for comparing the predictors or the empirical predictors, respectively. Finally, we identify (empirical) predictors with poor performance ($RPI > 50\%$) relatively to the best (empirical) predictor.

Since initial simulation results indicated that the magnitude of the (*EMSE*) *SMSE* is only slightly affected by changes in the shape of the response distribution, we confine our subsequent analysis to cases 1, 4 and 5 in Table 1. A preliminary analysis also showed that the empirical predictors for the *SP* model with shrinkage constant $\hat{k}_1^{(SP)}$ and for the *FM* model with shrinkage constant $\hat{k}_1^{(FM)}$ generally have lower *EMSE* than those for which the shrinkage constants are $\hat{k}_2^{(SP)}$ and $\hat{k}_2^{(FM)}$, respectively, so that in Sections 4.2 and 4.3 we restrict the analysis to the former. Due to space limitations, only a few tables and figures are presented for illustration purposes; a more extensive set of tables and figures are available at the web site (http://www.umass.edu/cluster/ed/).

[4] The *RPI* in A relative to B is defined as $RPI = \frac{A-B}{B} \times 100\%$.

Table 2
Maximum relative percent increase (*RPI*) in *SMSE* for the *CM*, *ME*, *SP* and *FM* model predictors relatively to the *SMSE* of the best predictor

Population	Model	Maximum relative percent increase in *SMSE*		
		$F = 0.2$	$F = 0.5$	$F = 0.8$
$N = 10, M = 5$	*CM*	104.472	410.709	702.816
	ME	36.433	62.335	68.598
	SP	64.231	259.148	447.711
	FM	0.026	0.009	0.003
$N = 10, M = 20$	*CM*	104.632	411.194	707.206
	ME	96.093	158.565	174.631
	SP	80.672	316.102	560.318
	FM	0.012	0.009	0.002
$N = 50, M = 20$	*CM*	907.801	2387.214	3861.398
	ME	181.850	194.393	197.749
	SP	713.619	1814.274	2966.400
	FM	0.003	0.001	0.001

4.1. Performance of predictors

The *FM* model predictor generally presents minimum *SMSE*. Only in a small number of settings (generally for extreme values (0.01 or 0.99) of ρ_s and ρ_t) the *ME* or the *SP* model predictors showed minimum *SMSE*. For these settings, the *RPI* in *SMSE* of the *FM* model predictor relatively to the best predictor is at most 0.03%, which may be justified by the variability introduced in the simulation process.

To complete the study of the performance of the different predictors, we consider a relative comparison of their *SMSE*. For the sampling fractions considered here, Table 2 shows the maximum *RPI* in *SMSE* for the *ME*, the *SP* and the *FM* model predictors with respect to the best predictor (generally obtained under the *FM* model). The *FM* model predictor is at least equivalent to the best predictor in all the settings. Excluding it, the *ME* model predictor is closer to the best predictor, having a lower maximum *RPI* in *SMSE* than either the *CM* or the *SP* model predictors. As the number of clusters and the cluster sampling fraction increase, the *RPI* in *SMSE* also increases.

In Table 3, we summarize general settings (depending on the cluster and unit intra-class correlation coefficients and overall cluster and unit sampling fractions), where the *CM*, the *ME* and the *SP* model predictors have similar *SMSE* to that of the best predictor (generally the *FM* model predictor).

From Table 3 we observe that all predictors have similar performance when both intra-class correlation coefficients are large ($\rho_s \geqslant 0.95$ and $\rho_t \geqslant 0.5$). The performance of the *ME* model predictor is more similar to that of the best predictor under a wider range of conditions (i.e., $\rho_s \geqslant 0.95$ or $\rho_t \leqslant 0.2$), followed by the *SP* model predictor when both intra-class correlation coefficients vary (from $\rho_s \geqslant 0.5$ and $\rho_t \geqslant 0.8$ to $\rho_s \geqslant 0.99$ and $\rho_t \geqslant 0.05$ with one correlation coefficient increasing as the other decreases).

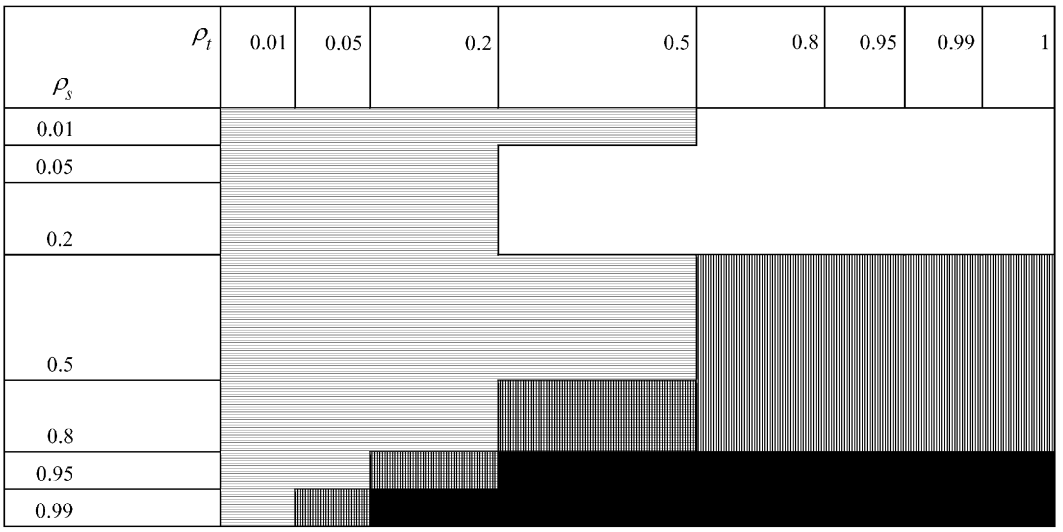
The predictor derived under the *ME* model has poor performance (i.e., $RPI > 50\%$) when the cluster intra-class correlation coefficient is small ($\rho_s \leqslant 0.2$), the unit intra-class correlation coefficient is large ($\rho_t \geqslant 0.95$) and the unit sampling fraction is large ($f \geqslant 0.8$) (see Fig. A.2 in the web site). Both the *SP* model predictor and the *CM* perform poorly as intra-class correlation coefficients tend jointly to zero, but this happens for increasing f in the case of the *SP* model predictor and for decreasing f in the case of the *CM* (see Figs. A.3 and A.1 in the web site). In contrast, the *FM* model predictor never presents poor performance.

4.2. Performance of empirical predictors

4.2.1. Loss of efficiency due to the use of empirical predictors

To evaluate the loss of efficiency associated to the use of empirical predictors instead of those where variance components are known, we computed the *RPI* of the *EMSE* with respect to the *SMSE* for each predictor. For the *ME* and the *FM* models, the *EMSE* associated with the empirical predictors is always greater than the *SMSE*, while for the

Table 3
Settings (depending on the intra-class correlation coefficients) where the *CM*, *ME* and *SP* model predictors have *SMSE* equivalent to that of the best predictor ($RPI < 5\%$)



Percent Increase in *MSE* relative to the best predictor. < 5% for the *ME* model. < 5% for the *SP* Model. < 5% for the *CM*.

Table 4
Descriptive statistics for the relative percent increase (*RPI*) in *EMSE* relatively to *SMSE* for each predictor

Population	Predictor	Min.	$Q1$	Mean	Median	$Q3$	Max.	St. Dev.
$N = 10$ $M = 5$	ME	0.1810	10.36	28.13	23.60	38.25	114.54	24.13
	SP	−5.1669	1.50	10.98	5.89	14.65	64.34	14.19
	FM	0.0259	6.76	24.05	21.18	34.72	80.10	20.15
$N = 10$ $M = 20$	*ME*	0.0384	5.81	23.01	18.51	31.55	163.34	22.57
	SP	−5.5243	0.70	12.83	4.51	16.36	119.45	20.07
	FM	0.0017	2.63	19.97	15.69	28.77	103.27	20.23
$N = 50$ $M = 20$	*ME*	0.0061	0.64	20.75	9.27	25.14	197.16	30.61
	SP	−10.9744	0.02	9.01	0.43	5.66	154.27	23.99
	FM	0.0002	0.30	16.56	4.00	19.68	141.35	25.83

SP model empirical predictor, the *EMSE* is greater than the *SMSE* in 85–96% of the settings. [5] In Table 4 we present some descriptive statistics for the *RPI* of the *EMSE* with respect to the *SMSE* for each predictor (see also Fig. A.4 in the web site for the corresponding box plots).

In general, the *SP* model empirical predictor presents a smaller efficiency loss than the other two empirical predictors, showing $|RPI| < 16\%$ in 75% of the settings, followed by the *FM* model empirical predictor and, lastly, by the *ME* model empirical predictor. These last two predictors show *RPI* lower than 35% and 38% in 75% of the settings, respectively.

[5] The settings where *EMSE* is smaller than *SMSE* generally correspond to situations where $MSB \leqslant MSR$ for almost all samples. When this happens, we may expect $\bar{Y}^*$ to be a better predictor than $\bar{Y}_i^*$. Additionally, when $MSB \leqslant MSR$, we generally have $k^{(SP)} \geqslant \hat{k}_1^{(SP)}$, so that the theoretical *SP* model predictor places more weight on $\bar{Y}_i^*$ (and less on $\bar{Y}^*$) than the empirical predictor, leading to the superiority of the latter.

Table 5
Percentage of settings where each predictor presents the best (minimum *EMSE*) or equivalent to the best ($0 < RPI < 15\%$) performance

Population	F	Model	% of cases where each predictor has minimum *EMSE*	% of cases where each predictor is "equivalent" to the best ($0 < RPI < 15\%$)	Total %
$N = 10, M = 5$	0.2	*CM*	**36.31**	16.07	52.38
		ME	16.67	59.52	76.19
		SP	17.86	55.36	73.22
		FM	**29.17**	67.26	**96.43**
	0.5	*CM*	15.48	29.76	45.24
		ME	19.05	63.69	82.74
		SP	**38.69**	19.64	58.33
		FM	**26.79**	63.69	**90.48**
	0.8	*CM*	0	44.05	44.05
		ME	19.05	59.52	78.57
		SP	**41.07***	16.07	57.14
		FM	**41.67***	48.81	**90.48**
$N = 10, M = 20$	0.2	*CM*	**46.94**	14.54	61.48
		ME	18.11	65.05	83.16
		SP	16.07	72.19	88.26
		FM	**18.88**	81.12	**100.00**
	0.5	*CM*	14.80	40.82	55.62
		ME	17.09	70.15	87.24
		SP	**48.47***	27.81	76.28
		FM	**20.41***	76.28	**96.69**
	0.8	*CM*	0.51	52.30	52.81
		ME	17.09	69.64	86.73
		SP	**51.79***	22.19	73.98
		FM	**31.89***	63.52	**95.41**
$N = 50, M = 20$	0.2	*CM*	0	49.23	49.23
		ME	17.09	68.88	85.97
		SP	**45.66***	26.53	72.19
		FM	**38.01***	57.65	**95.66**
	0.5	*CM*	0	45.41	45.41
		ME	15.56	70.41	85.97
		SP	**35.97***	32.65	68.62
		FM	**50.26***	45.92	**96.18**
	0.8	*CM*	0	45.41	45.41
		ME	15.82	69.90	85.72
		SP	**30.10***	37.50	67.60
		FM	**56.12***	40.05	**96.17**

The worst performance for all predictors (not shown) is attained when unit-sampling fractions are small, especially when both cluster and unit intra-class correlation coefficients decrease. An exception occurs for the *ME* model empirical predictor, where this poor performance is also observed for large unit sampling fractions when $\rho_t \geqslant 0.8$ and for varying values of ρ_s depending on population and cluster sampling fractions.

4.2.2. Comparison of the EMSE

We computed the percentage of settings where each empirical predictor satisfies each of the three adopted criteria, namely: (a) minimum *EMSE*, (b) "equivalence" to the best empirical predictor (i.e., $RPI < 15\%$) and (c) poor performance ($RPI > 50\%$). Although these percentages depend on the selection of settings, i.e., on the specified population sizes and sampling plans considered in this investigation, such a summary is one way to provide an overall description of results. All the percentages are computed considering the combination of all intra-class correlation coefficients and unit sampling fractions in the denominator, i.e., 168($=7 \times 8 \times 3$) for the populations with $N = 10$ and $M = 5$ and

392(=7 × 8 × 7) for the other populations (Tables 5 and 7). In some cases (identified by * in Table 5), the sums of the percentages exceed 100% because the *EMSE* for the *SP* and *FM* model empirical predictors have exactly the same minimum value. In Table 5, the boldfaced figures identify the percentage of settings for the two empirical predictors with the best performance for each type of population and cluster-sampling fraction. The *FM* model empirical predictor does not always have the best performance in terms of the minimum *EMSE*.

For populations with $N = 10$, the *CM* presents the minimum *EMSE* for small cluster sampling fractions ($F = 0.2$), followed by the *FM* model empirical predictor. As the cluster sampling fraction increases, the *FM* or the *SP* model empirical predictors are best. For populations with $N = 50$, the *SP* or the *FM* model empirical predictors also appear as the two best ones, with the first being better for small cluster sampling fractions ($F = 0.2$) and the second, for moderate to large sampling fractions ($F = 0.5, 0.8$).

We identify certain patterns (depending on ρ_s and ρ_t) where each empirical predictor can be considered equivalent (in terms of *EMSE*) to the best one. Table 6 summarizes these results ($RPI < 15\%$) for all cluster and unit sampling

Table 6
Settings (depending on the intra-class correlation coefficients) where the *CM*, *ME*, *SP* and the *FM* empirical predictors have similar performance ($RPI < 15\%$) relatively to the predictor with minimum *EMSE*

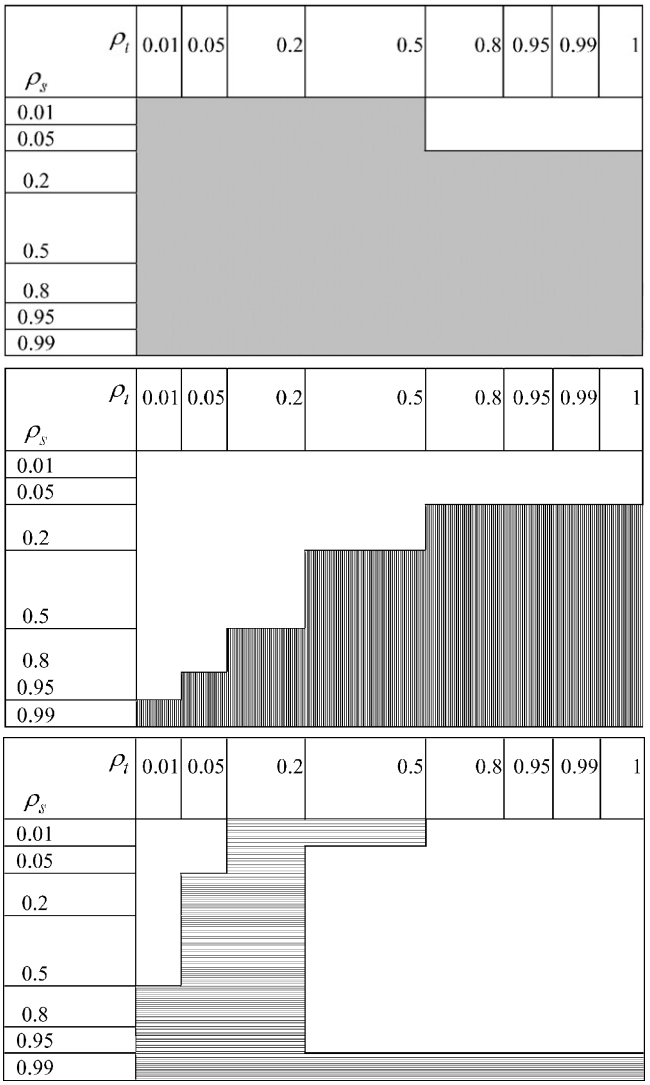
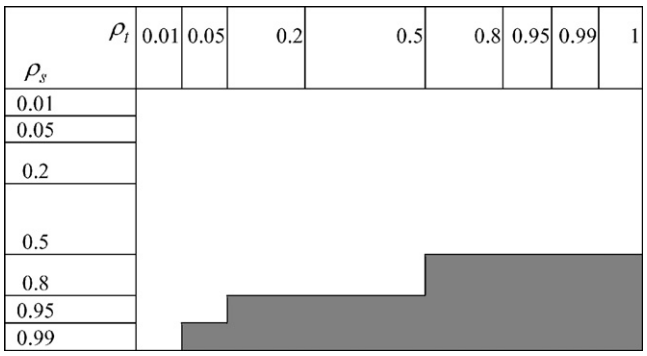


Table 6 (*contd.*)



Percent increase in *MSE* relative to best predictor. < 15% for the *FM* model predictor. < 15% for the *SP* model predictor. < 15% for the *ME* model predictor. < 15% for the *CM*.

Table 7
Poor performance of each empirical predictor

Population	F	Model	% of cases where each predictor has poor performance relative to the best ($RPI > 50\%$)	Maximum *RPI* (%)
$N = 10, M = 5$	0.2	*CM*	5.95	67.76
		ME	4.17	79.55
	0.5	*CM*	43.45	273.50
		ME	5.95	192.82
		SP	26.19	177.11
	0.8	*CM*	47.02	473.70
		ME	6.55	245.37
		SP	32.14	305.84
$N = 10, M = 20$	0.2	*CM*	5.10	53.68
		ME	4.34	165.14
	0.5	*CM*	35.46	225.10
		ME	4.85	468.41
		SP	10.97	186.19
	0.8	*CM*	38.27	392.25
		ME	5.36	608.48
		SP	14.29	321.28
$N = 50, M = 20$	0.2	*CM*	39.29	500.27
		ME	5.61	574.51
		SP	14.54	414.36
	0.5	*CM*	41.58	1284.02
		ME	6.63	718.50
		SP	18.62	1046.82
	0.8	*CM*	42.35	2038.30
		ME	6.89	780.11
		SP	20.15	1676.37

fractions, but these regions may be extended for specific numbers and sizes of clusters and cluster or unit sampling fractions.[6]

Independently of the population characteristics and sampling fractions, the *FM* model empirical predictor is the best or equivalent to the best in a larger number of settings (90–100%) than any competitor (see Table 5).

As in the case of known variances, both the *CM* and the empirical predictor derived under the *SP* model have poor performance ($RPI > 50\%$) in many settings (up to 47% and 32%, respectively), followed by the *ME* empirical predictor (up to 7%). These results may be visualized in Table 7.

In particular, the empirical predictors derived under the *ME* model show a poor performance when the cluster intra-class correlation coefficient is small ($\rho_s \leqslant 0.5$), the unit intra-class correlation coefficient is large ($\rho_t \geqslant 0.8$) and the unit sampling fraction is large ($f \geqslant 0.6$). For these empirical predictors, the maximum overall *RPI* is 780%. The *SP* model empirical predictor and the *CM* perform poorly as the intra-class correlation coefficients tend jointly to zero, but this happens for increasing f in the case of the *SP* model empirical predictor and for decreasing f in the case of the *CM*. The *RPI* reaches an overall maximum value of 1676% for the *SP* model empirical predictors and 2038% for the *CM*. In contrast, the *FM* model empirical predictor never has such a poor performance. This suggests that, in practice, this empirical predictor is more robust (in the sense of not having very bad performance) than its competitors.

5. Discussion

The *FM* model empirical predictor shows a more stable performance than its competitors, being the best or equivalent to the best empirical predictor in 90–100% of the settings; furthermore, it never exhibits a poor performance. The response distributions under consideration have almost no effect on the values of the *MSE*.

In the absence of response error and when all the units in each sampled cluster are observed, both the *SP* and the *FM* model empirical predictors reproduce the cluster mean, while the *ME* model empirical predictor does not. This highlights the focus of the *ME* model on predicting a conceptual latent value where the observed portion (the *PSU* sample data) has an insignificant contribution to the predictor since *PSU*s are assumed to contain infinite units.

The superiority of the *SP* model empirical predictor under some of the settings considered in this investigation may be attributed to a smaller loss in efficiency for the empirical *SP* predictor than for the empirical *FM* predictor. Nevertheless, the performance of the empirical predictor obtained under the *FM* model improves as the number of clusters and the cluster sampling fraction increase, becoming the best over a wider range of settings.

We based the simulation studies on the actual physical process that underlies sampling design and measurement. This approach accounts simultaneously for lack of fit, response error and the two-stage sampling. Our conclusions are restricted to the case of identical within-cluster variances. We anticipate that the results presented here would be similar to results from simulation studies based on an extended superpopulation model that includes lack of fit and response error.

Another study considering different within-cluster variances is in progress and preliminary results are consistent with the present conclusions. Keeping the limitations of simulation studies in mind, our results point in the direction of recommending the *FM* model empirical predictor against its competitors in a variety of settings. Further developments to obtain the *MSE* (or some approximation) for the empirical predictors under the *FM* model are required for practical applications.

Acknowledgments

This work was developed with financial support from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil and the National Intitutes of Health (NIH-PHS-R01-HD36848, R01-HL071828-02), USA. The authors are grateful to the anonymous referees for their enlightening comments.

[6] The 'white' area in Table 6 (for $\rho_s \leqslant 0.05$ and $\rho_t \geqslant 0.8$) indicates that there is no best predictor over all cluster and unit sampling fractions.

References

Bolfarine, H., Zacks, S., 1992. Prediction Theory for Finite Populations. Springer, New York.
Goldberger, A.S., 1962. Best linear unbiased prediction in the generalized linear regression model. J. Amer. Statist. Assoc. 57, 369–375.
Hedayat, A.S., Sinha, B.K., 1991. Design and Inference in Finite Population Sampling. Wiley, New York.
Henderson, C.R., 1984. Applications of Linear Models in Animal Breeding. Mimeo, University of Guelph, Guelph, Canada.
Kackar, R.N., Harville, D.A., 1984. Approximations of standard errors of estimators for fixed and random effects in mixed linear models. J. Amer. Statist. Assoc. 79, 853–863.
Li, Y.D., Lahiri, P., 2006. Robust model-based and model-assisted predictors of the finite population total. J. Amer. Statist. Assoc. 102, 664–673.
McCulloch, C.E., Searle, S.R., 2001. Generalized, Linear and Mixed Effects Models. Wiley, New York.
McLean, R.A., Sanders, W.L., Stroup, W.W., 1991. A unified approach to mixed linear models. Amer. Statist. 45, 54–64.
Moura, F.A.S., Holt, D., 1999. Small area estimation using multilevel models. Survey Methodology 25, 73–80.
Pfeffermann, D., Nathan, G., 1981. Regression analysis of data from a cluster sample. J. Amer. Statist. Assoc. 76, 681–689.
Prasad, N.G.N., Rao, J.N.K., 1990. The estimation of the mean square error of small-area estimators. J. Amer. Statist. Assoc. 85, 163–171.
Robinson, G.K., 1991. That BLUP is a good thing: the estimation of random effects. Statist. Sci. 6, 15–51.
Scott, A., Smith, T.M.F., 1969. Estimation in multi-stage surveys. J. Amer. Statist. Assoc. 64, 830–840.
Searle, S.R., Fawcett, R.F., 1970. Expected mean squares in variance components models having finite populations. Biometrics 26, 243–254.
Searle, S.R., Casella, G., McCulloch, C.E., 1992. Variance Components. Wiley, New York.
Stanek III, E.J., Singer, J.M., 2004. Predicting random effects from finite population clustered samples with response error. J. Amer. Statist. Assoc. 99, 1119–1130.
Stanek III, E.J., Singer, J.M., Lencina, V.B., 2004. A unified approach to estimation and prediction under simple random sampling. J. Statist. Plann. Inference 121, 325–338.
Stanek, E.J., Well, A., Ockene, I., 1999. Why not routinely use best linear unbiased predictors (BLUPs) as estimates of cholesterol, per cent fat from Kcal and physical activity. Statist. Med. 18, 2943–2959.