Performance of Balanced Two-Stage Empirical Predictors of Realized Cluster Latent Values from Finite Populations: A Simulation Study

Silvina San Martino[1], *Julio da Motta Singer*[2], *Edward J. Stanek III*[3]

Abstract

Predictors of random effects are usually based on the popular mixed effects model developed under the assumption that the sample is obtained from a conceptual infinite population even when the actual population is finite. Two alternatives that incorporate the finite nature of the population are the super-population model proposed by Scott and Smith (1969, JASA, 64: 830-840) and the random permutation model recently proposed by Stanek and Singer (2004, JASA, 99:1119-1130). The random permutation model based predictor derived under the additional assumptions that all variance components are known and that within cluster variances are equal has smaller mean squared error than the

[1] Facultad de Ciencias Agrarias, Universidad Nacional de Mar del Plata, Mar del Plata, Argentina. ssanmarti@balcarce.inta.gov.ar.

[2] Departamento de Estatística, Universidade de São Paulo, São Paulo, Brazil. jmsinger@ime.usp.br.

[3] Department of Public Health, University of Massachusetts, Amherst, MA, USA. stanek@schoolph.umass.edu.

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

corresponding predictors based on either the mixed effects 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 simulation results suggest that the performance of the random permutation model empirical predictor improves either as the cluster sampling fractions increase or as the number and size of clusters increase. 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. Additionally, the results indicate that the random permutation model empirical predictor is more stable than its competitors since, in terms of mean squared error, it is either the best or the second best and even in this case, its performance lies within acceptable limits, and never presents a very poor behavior.

Keywords: Empirical predictors, Finite population; Optimal prediction;; Random permutation; Two-stage sampling

Deleted: and never presents a very poor behavior.

Deleted: Two-stage sampling

Deleted: Optimal prediction; Empirical predictors.

Deleted: SanMartino&al2005v20.doc

Deleted: 18/01/07

Deleted: 09/01/07

Inserted: SanMartino&al2005v20.doc

Deleted: SanMartino&al2005v17.doc

Inserted: 18/01/07

Deleted: 14/02/07

Inserted: 14/02/07

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, usually, there is interest in the evaluation of the contribution of within and between cluster variability to overall variability with information extracted from a multi-stage random sample selected without replacement. In particular, when interest lies in the prediction of the latent values of realized clusters (i.e., the average expected response of the units in those clusters) based on data from a two-stage sample from a finite population, three approaches may be considered.

The most popular approach is based on the usual mixed model derived under the assumption that the sample is obtained from a conceptual infinite population. Best linear unbiased predictors (BLUP) derived under such models have been widely considered in the literature (Goldberger (1962), Henderson (1984), McLean, Sanders, and Stroup (1991), Robinson (1991), Stanek, Well, and Ockene (1999), McCulloch and Searle (2001)), but they do not account for the finite nature of the population under investigation. 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 super-population model. This approach has had limited application, in part because its performance may be severely affected by model miss-specification and because it relies on the artificial nature of the postulated super-population. The third, recently suggested by Stanek, Singer and Lencina (2003) in a simple random sampling setup and

Deleted: Although authors like Searle, Casella and McCulloch (1992) or Cox and Solomon (2003) mention that finite populations arise quite rarely in applications, there

Deleted: . This is true

Deleted: due to its sensitivity to model

Deleted: to the

Deleted: SanMartino&al2005v20.doc

Deleted: 18/01/07

Deleted: 09/01/07

Inserted: SanMartino&al2005v20.doc

Deleted: SanMartino&al2005v17.doc

Inserted: 18/01/07

Deleted: 14/02/07

Inserted: 14/02/07

extended by Stanek and Singer (2004) to a balanced two-stage sampling setup with or without response error, considers a design-based probability model, (called "random permutation model" (*RP*)), induced by the sampling process. This approach does not require restrictive assumptions (such as normality) so that it can be applied to a wide range of practical settings.

Under each approach, the predictors of realized cluster latent values are formed as the sum of predictors of observed and unobserved units, differing with respect to the associated shrinkage factors. These shrinkage factors are functions of population variance components and, for the last two models, they depend on finite population characteristics such as cluster sizes and on the sampling fraction of units. Depending on all these characteristics, 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 with respect to a certain subject by means of a test with scores ranging from 0 to 10. We assume that the student responses include measurement 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 a sample of classrooms in the school. In addition to estimating the school response and variance components, suppose that there is interest in predicting response for a sampled classroom. For illustration, let us assume that the between classroom variability is 1.25, the within cluster variability is 2.00 and that the response error variability is 0.80 (i.e., the cluster intra-class correlation is 0.38

Deleted: the

Deleted: Let

Deleted: for illustration

Deleted: SanMartino&al2005v20.doc

Deleted: 18/01/07

Deleted: 09/01/07

Inserted: SanMartino&al2005v20.doc

Deleted: SanMartino&al2005v17.doc

Inserted: 18/01/07

Deleted: 14/02/07

Inserted: 14/02/07

and the unit intra-class correlation is 0.71). Suppose that the school sample average is 6.75, while for the classroom with teacher i, the sample average is 5.20. Based on the sample data, there are four approaches to predicting the latent classroom mean. First, we may use the sample classroom average, 5.20. Alternatively, assuming that the response error model holds for all students, the average performance for the classroom with teacher *i* is predicted to be 5.40, 5.30 and 5.90 respectively using the mixed effects model, Scott and Smith's model or Stanek and Singer´s *RP* model predictors (see section 2.4 for details about how each predictor is obtained). The 11% relative difference observed between the predicted values obtained under the random permutation model and Scott and Smith's 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. The objective of this comparison is to select the predictor that provides the best result, using the mean squared error (*MSE*) as the criterion for selecting the best predictor.

The mixed effects, Scott and Smith's and random permutation models rely on different assumptions. Only the *RP* model links the finite population to a set of random variables without requiring additional assumptions than those related to simple random sampling. 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

Deleted: 4

Deleted: aim

Deleted: most desirable

Deleted: a

Deleted: criteria

Deleted: is based on the sampling scheme and directly

Deleted: with

Deleted: SanMartino&al2005v20.doc

Deleted: 18/01/07

Deleted: 09/01/07

Inserted: SanMartino&al2005v20.doc

Deleted: SanMartino&al2005v17.doc

Inserted: 18/01/07

Deleted: 14/02/07

Inserted: 14/02/07

model have smaller *MSE* (evaluated under the *RP* model assumptions) than those based on the mixed effects and Scott and Smith's models. Here we consider simulation studies to compare the predictors in terms of the empirical *MSE*.

In practical situations, variances are rarely known and need to be estimated. In this context, 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 three 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/Results-pub.html.

2. Predictors 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

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

Deleted: random permutation

Deleted: From a practical point of view, it may be more realistic to

Deleted: without assuming a specific covariance structure

Deleted: SanMartino&al2005v20.doc

Deleted: 18/01/07

Deleted: 09/01/07

Inserted: SanMartino&al2005v20.doc

Deleted: SanMartino&al2005v17.doc

Inserted: 18/01/07

Deleted: 14/02/07

Inserted: 14/02/07

We define the latent value in cluster s as $\mu_s = \frac{1}{M}\sum_{t=1}^{M} y_{st}$ and the corresponding variance as $\left(\frac{M-1}{M}\right)\sigma_s^2 = \frac{1}{M}\sum_{t=1}^{M}\left(y_{st}-\mu_s\right)^2$ for $s = 1,...,N$. Also, we let $\sigma_e^2 = \frac{1}{N}\sum_{s=1}^{N}\sigma_s^2$ denote the average within cluster variance. Similarly, we define the population mean and the between cluster variance as $\mu = \frac{1}{N}\sum_{s=1}^{N}\mu_s$ and $\left(\frac{N-1}{N}\right)\sigma^2 = \frac{1}{N}\sum_{s=1}^{N}\left(\mu_s-\mu\right)^2$, respectively.

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 random permutation (*RP*) model

We briefly summarize the *RP* model presented by Stanek and Singer (2004). This model is induced by the two-stage random sampling of a finite clustered population. The two-stage random permutation model is represented as an ordered list of NM random variables, the values of which are the responses of an independent permutation of clusters and units in clusters. For each permutation, we assign a new label, $i = 1,...,N$ 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 $j = 1,...,M$. For ease of exposition, we refer to the cluster that will occupy position i in the permutation of clusters as primary sampling unit

Deleted: that

Deleted: SanMartino&al2005v20.doc

Deleted: 18/01/07

Deleted: 09/01/07

Inserted: SanMartino&al2005v20.doc

Deleted: SanMartino&al2005v17.doc

Inserted: 18/01/07

Deleted: 14/02/07

Inserted: 14/02/07

(*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 indicator random variables: U_{is}, which takes on a value of one when the realized cluster corresponding to *PSU* i is cluster s and a value of zero otherwise, and $U_{jt}^{(s)}$, which takes 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 *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 the realization of the random variable

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

where $\mathbf{Y} = \begin{pmatrix}\mathbf{Y}_1' & \mathbf{Y}_2' & \cdots & \mathbf{Y}_N'\end{pmatrix}' \in \mathbb{R}^{NM}$, with $\mathbf{Y}_i = \begin{pmatrix}Y_{i1} & Y_{i2} & \cdots & Y_{iM}\end{pmatrix}' \in \mathbb{R}^{M}$,

$\mathbf{U}^{(s)} = \begin{pmatrix}\mathbf{U}_1^{(s)} & \mathbf{U}_2^{(s)} & \cdots & \mathbf{U}_M^{(s)}\end{pmatrix} \in \mathbb{R}^{M}\times\mathbb{R}^{M}$, with $\mathbf{U}_t^{(s)} = \begin{pmatrix}U_{1t}^{(s)} & U_{2t}^{(s)} & \cdots & U_{Mt}^{(s)}\end{pmatrix}'$, and

$\mathbf{U} = \begin{pmatrix}\mathbf{U}_1 & \mathbf{U}_2 & \cdots & \mathbf{U}_N\end{pmatrix} \in \mathbb{R}^{N}\times\mathbb{R}^{N}$, with columns $\mathbf{U}_s = \begin{pmatrix}U_{1s} & U_{2s} & \cdots & U_{Ns}\end{pmatrix}'$.

Here, $\otimes$ denotes the Kronecker product and $\bigoplus_{s=1}^{N}\mathbf{A}_s$ denotes a block diagonal matrix with blocks $\mathbf{A}_s$ (Searle, Casella and McCulloch, 1992). The difference between $\mathbf{Y}$ and $\mathbf{y}$ is the interpretation of the subscripts that define the elements of the vectors. The subscripts in $\mathbf{Y}$

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

correspond to positions in a permutation, while the subscripts in $\mathbf{y}$ correspond to labels of units.

Defining $\beta_s = (\mu_s - \mu)$ as the deviation of the latent value of cluster *s* from the population mean and $\varepsilon_{st} = (y_{st} - \mu_s)$ as 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

$$\mathbf{y} = \mathbf{X}\mu + \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (2.1)$$

where $\mathbf{X} = \mathbf{1}_N \otimes \mathbf{1}_M$, $\mathbf{Z} = \mathbf{I}_N \otimes \mathbf{1}_M$, $\boldsymbol{\beta}' = (\beta_1 \quad \beta_2 \quad \cdots \quad \beta_N)$, 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 1.

Using elementary properties of the indicator random variables and the simple structure of $\mathbf{X}$ and $\mathbf{Z}$, it follows that $\mathbf{U}\mathbf{1}_N = \mathbf{1}_N$, $\mathbf{U}^{(s)}\mathbf{1}_M = \mathbf{1}_M$, $(\mathbf{U} \otimes \mathbf{I}_M)\left(\bigoplus_{s=1}^{N} \mathbf{U}^{(s)}\right)\mathbf{X} = \mathbf{X}$ and $(\mathbf{U} \otimes \mathbf{I}_M)\left(\bigoplus_{s=1}^{N} \mathbf{U}^{(s)}\right)\mathbf{Z} = \mathbf{U} \otimes \mathbf{1}_M = \mathbf{Z}\mathbf{U}$. 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 random permutation mixed effects model

$$\mathbf{Y} = \mathbf{X}\mu + \mathbf{Z}\mathbf{B} + \mathbf{E} \qquad (2.2)$$

where $\mathbf{E} = (\mathbf{U} \otimes \mathbf{I}_M)\left(\bigoplus_{s=1}^{N} \mathbf{U}^{(s)}\right)\boldsymbol{\varepsilon}$ and $\mathbf{B} = \mathbf{U}\boldsymbol{\beta} = (B_1 \quad B_2 \quad \cdots \quad B_N)'$. Note that because of the random variables $\mathbf{U}$, the terms $B_i = \sum_{s=1}^{N} U_{is}\beta_s$ for $i = 1, \ldots, N$ are random effects and represent the deviation of the latent value for *PSU i* from the population mean.

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

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

As the values of the finite population defined by $\mathbf{y}$ may not be observed directly, we assume a response error 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 $\mathrm{var}(W_{st}) = \sigma_{st}^2$, $s = 1,\ldots,N$, $t = 1,\ldots,M$.

Using this notation, the two-stage random permutation model with response error is represented by

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

where $\mathbf{W}^* = (\mathbf{U} \otimes \mathbf{I}_M)\left(\bigoplus_{s=1}^{N}\mathbf{U}^{(s)}\right)\mathbf{W}$. Under the re-parameterization (2.1), we express this as a mixed effects model of the form

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

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

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

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

and

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

The subscript ξ_3 denotes expectation with respect to response error and $\sigma_r^2=\sum_{s=1}^{N}\sum_{t=1}^{M}\frac{\sigma_{st}^2}{NM}$ 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 (i.e., permutation) has been selected, it will be apparent which cluster corresponds to a particular *PSU*. We refer to that cluster as the **realized *PSU*.**

Denoting the sample elements by $\mathbf{Y}_I^*=\mathbf{Y}_I+\mathbf{W}_I^*$, the model for the sample is $\mathbf{Y}_I^*=\mathbf{X}_I\mu+\mathbf{Z}_I\mathbf{B}+\left(\mathbf{E}_I+\mathbf{W}_I^*\right)$ where $\mathbf{X}_I=\mathbf{1}_n\otimes\mathbf{1}_m$, $\mathbf{Z}_I=\mathbf{I}_n\otimes\mathbf{1}_m$, $E\left(\mathbf{Y}_I^*\right)=\mathbf{X}_I\mu$ and

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

2.2. Scott and Smith´s (*SS*) super-population model

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

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

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

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 directly linked to the population units as in the random permutation 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 super-population such that for each $i=1,\ldots,N$, $E(\Lambda_i)=\mu$ and $\text{var}(\Lambda_i)=\sigma^2$. Then, σ^2 may be interpreted as the variance of the distributions of the random variables $\Lambda_1,\ldots,\Lambda_N$ from which the cluster means $\mu_1,\ldots,\mu_N$ constitute a sample. According to this model, elements within the same cluster are correlated, but elements in different clusters are not.

Using Bayesian models, Bolfarine and Zacks (1992) extend the approach considered by Scott and Smith (1969) to a two-stage sampling process with response error. Essentially, they add a normality assumption to the super-population model considered by Scott and Smith (1969) and suppose that the potentially observed variables are given by elements of

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

where $\mathbf{W} \sim N\left(\mathbf{0}, \bigoplus_{i=1}^{N} \sigma_{ri}^2 \mathbf{I}_M\right)$, $i=1,...,N$, $j=1,...,M$, and $\mathbf{W}$ is independent of $\mathbf{Y}$.

Assuming $\sigma_{ri}^2 = \sigma_r^2$, $i=1,...,N$, we obtain

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

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

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

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

2.3. The mixed effects (*ME*) model

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

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

where $\mathbf{X}_{Ii} = \mathbf{Z}_{Ii} = \mathbf{1}_m$ and $\mathbf{E}_{Ii} = \left(E_{i1}\, E_{i2} \ldots E_{im}\right)'$. 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 $B_i \sim N\left(0, \sigma^2\right)$ and $\mathbf{E}_{Ii} \sim N\left(\mathbf{0}, \sigma_i^2\mathbf{I}_m\right)$ are independent, so that $\mathbf{Y}_{Ii} \sim N\left(\mathbf{X}_{Ii}\mu,\, \sigma_i^2\mathbf{I}_m + \sigma^2\mathbf{J}_m\right)$. If in (2.6) we include a second additive (response) error term $\mathbf{W}_{Ii} \sim N\left(\mathbf{0}, \sigma_r^2\mathbf{I}_m\right)$ independent of B_i and $\mathbf{E}_{Ii}$ then $\mathbf{Y}_{Ii}^* = \mathbf{Y}_{Ii} + \mathbf{W}_{Ii} \sim N\left(\mathbf{X}_{Ii}\mu, \left(\sigma_i^2 + \sigma_r^2\right)\mathbf{I}_m + \sigma^2\mathbf{J}_m\right)$. It follows that

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

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

where $\mathbf{B}_I = \begin{pmatrix} B_1 & B_2 & \cdots & B_n \end{pmatrix}'$, $\mathbf{E}_I = \begin{pmatrix} \mathbf{E}'_{I1} & \mathbf{E}'_{I2} & \cdots & \mathbf{E}'_{In} \end{pmatrix}'$, and

$\mathbf{W}_I = \begin{pmatrix} \mathbf{W}'_{I1} & \mathbf{W}'_{I2} & \cdots & \mathbf{W}'_{In} \end{pmatrix}'$ which implies that $\mathbf{Y}_I^* \sim N\left(\mathbf{X}_I \mu, \bigoplus_{i=1}^{n} \left[\left(\sigma_i^2 + \sigma_r^2 \right) \mathbf{I}_m + \sigma^2 \mathbf{J}_m \right] \right)$.

Here again, neither μ nor σ^2 and σ_i^2 refer to the finite population mean or variance components. Instead, the variance of the random effects, σ^2, can be interpreted as the variance of the random cluster mean $\left(\mu + B_i \right)$ which conceptually takes on an infinite number of values.

2.4. Predictors 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 $i \le n$), i.e., $T_i = \mathbf{g}'\mathbf{Y}$, where $\mathbf{g}' = \frac{1}{M} \mathbf{e}_i' \otimes \mathbf{1}_M'$, and $\mathbf{e}_i$ denotes an $N \times 1$ column vector with 1 in position i and zero elsewhere. In the mixed model, T_i corresponds to $\mu + B_i$. From the sampled values $\mathbf{Y}_I^* = \begin{pmatrix} \mathbf{Y}_{I1}^{*\prime} & \mathbf{Y}_{I2}^{*\prime} & \ldots & \mathbf{Y}_{In}^{*\prime} \end{pmatrix}'$, where $\mathbf{Y}_{Ii}^* = \begin{pmatrix} Y_{i1}^* & Y_{i2}^* & \ldots & Y_{im}^* \end{pmatrix}'$, the predictors of T_i (for $i \le n$) under the three models (i.e. (2.3), (2.5) or (2.7)) may be obtained as follows:

i) For the mixed effects model: $\hat{T}_i = \hat{\mu} + k_i^{(ME)} \left(\bar{Y}_i^* - \hat{\mu} \right)$,

where $\bar{Y}_i^* = \frac{1}{m} \sum_{j=1}^{m} Y_{ij}^*$, $\hat{\mu} = \frac{\sum_{i=1}^{n} \bar{Y}_i^* / \nu_i}{\sum_{i=1}^{n} 1/\nu_i}$, $\nu_i = m\sigma^2 + \sigma_i^2 + \sigma_r^2$, and the shrinkage constant is

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

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

ii) For the Scott and Smith model: $\hat{T}_i = \hat{\mu} + k_i^{(SS)}\left(\bar{Y}_i^* - \hat{\mu}\right)$,

where the shrinkage constant is

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

with $f = \frac{m}{M}$ denoting the sampling fraction for units.

iii) For the random permutation model: $\hat{T}_i = \bar{Y}^* + k^{(RP)}\left(\bar{Y}_i^* - \bar{Y}^*\right)$,

where $\bar{Y}^* = \frac{1}{n}\sum_{i=1}^{n}\bar{Y}_i^*$ and the shrinkage constant is

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

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

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

where

i) For the mixed effects model: $k^{(ME)} = \frac{m\sigma^2}{m\sigma^2 + \sigma_e^2 + \sigma_r^2}$.

ii) For the Scott and Smith model:

$$k^{(SS)} = k^{(ME)} + f\left(1 - k^{(ME)}\right) = \frac{m\sigma^2 + f\left(\sigma_e^2 + \sigma_r^2\right)}{m\sigma^2 + \sigma_e^2 + \sigma_r^2}.$$

iii) For the random permutation model:

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

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

In this case, the shrinkage constants are such that $0 \le k^{(ME)} \le k^{(RP)} \le k^{(SS)} \le 1$.

The study of 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 Figure 1, we compare the behavior of the theoretical shrinkage constants as *f*, and the cluster (ρ_s) and unit (ρ_t) intra-class correlation coefficients[5] vary. We also evaluate the performance of the cluster mean (*CM*) for which the shrinkage constant $k^{(CM)}$ is always equal to 1. The green solid line identifies the theoretical shrinkage constant for the *RP* model predictor. Since the random permutation model predictors have smallest *MSE* when variance components are known, we consider it as reference to compare the predictors and identify similar as well as different (poor) performance relative to it.

Insert Figure 1 here

We expect minor differences between the predictors, almost independently of *f*, when both ρ_s and ρ_t tend to one. From Figure 1, we also observe that when f tends to 0,

[5] We define the cluster intra-class correlation coefficient as $\rho_s = \frac{\sigma^2}{\sigma^2 + \sigma_e^2}$ and the unit intra-class correlation coefficient as $\rho_t = \frac{\sigma_e^2}{\sigma_e^2 + \sigma_r^2}$.

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

$k^{(ME)}$, $k^{(SS)}$ and $k^{(RP)}$ 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), $k^{(RP)} = k^{(SS)} = k^{(CM)} = 1$ (and all differ from $k^{(ME)}$), so that the best predictor is the cluster mean. When ρ_t tends to zero, $k^{(ME)}$ approaches $k^{(RP)}$ (almost independently of f), so that for this situation, we expect the predictors derived under these two models to behave similarly.

Deleted: that

Relative to the theoretical results for the *RP* 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 *SS* 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.

2.5. Empirical predictors 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 substituting the shrinkage constants by their respective estimators. Searle and Fawcett (1970) developed a rule for converting expectations of mean squares obtained under variance component infinite population models into expectations under finite population models to estimate variance components, 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

Deleted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

structure of the first two moments, we use method of moments estimators of variance components. These estimators may be derived from ANOVA mean squares, namely $MSB = (n-1)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\bar{Y}_i^* - \bar{Y}^*\right)^2$ and $MSR = [n(m-1)]^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(Y_{ij}^* - \bar{Y}_i^*\right)^2$. The mean squares may be expressed as quadratic forms of the type $\mathbf{Y}_I^{*\prime}\mathbf{A}\mathbf{Y}_I^*$ where $\mathbf{A} = (n-1)^{-1}\left(\mathbf{P}_n \otimes \frac{\mathbf{J}_m}{m}\right)$ for *MSB* or $\mathbf{A} = [n(m-1)]^{-1}\left(\mathbf{I}_n \otimes \mathbf{P}_m\right)$ for *MSR*, with $\mathbf{P}_a = \mathbf{I}_a - a^{-1}\mathbf{J}_a$ and a denoting a positive integer. Since $\mathbf{A}\mathbf{1}_{nm} = \mathbf{0}$ and $E\left(\mathbf{Y}_I^*\right) = \mathbf{1}_{nm}\mu$, under any of the three models, it follows that $E\left(\mathbf{Y}_I^{*\prime}\right)\mathbf{A}E\left(\mathbf{Y}_I^*\right) = 0$. Therefore,

$$E\left(\mathbf{Y}_I^{*\prime}\mathbf{A}\mathbf{Y}_I^*\right) = tr\left[\operatorname{var}\left(\mathbf{Y}_I^*\right)\mathbf{A}\right] + E\left(\mathbf{Y}_I^{*\prime}\right)\mathbf{A}E\left(\mathbf{Y}_I^*\right) = tr\left[\operatorname{var}\left(\mathbf{Y}_I^*\right)\mathbf{A}\right].$$

Using these results, we obtain method of moments 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. *RP* model

Under the random permutation model, we have

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

To evaluate the expected value of *MSR*, we let $\mathbf{A} = [n(m-1)]^{-1}\left(\mathbf{I}_n \otimes \mathbf{P}_m\right)$, so that

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

$$\text{var}\left(\mathbf{Y}_I^*\right)\mathbf{A} = \frac{\left(\sigma_e^2 + \sigma_r^2\right)}{n\left(m-1\right)}\left(\mathbf{I}_n \otimes \mathbf{P}_m\right),$$

which implies that

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

To evaluate the expected value of *MSB*, we let $\mathbf{A} = (n-1)^{-1}\left(\mathbf{P}_n \otimes \frac{\mathbf{J}_m}{m}\right)$, so that

$$\text{var}\left(\mathbf{Y}_I^*\right)\mathbf{A} = \frac{\sigma_e^2 + \sigma_r^2}{n-1}\left(\mathbf{P}_n \otimes \frac{\mathbf{J}_m}{m}\right) + \frac{\sigma^{*2}}{n-1}\mathbf{P}_n \otimes \mathbf{J}_m$$

which implies that

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

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 $m\hat{\sigma}^2 = MSB - \left(1-f\right)MSR - f\sigma_r^2$. Consequently, one method of moments estimator for $k^{(RP)}$ is

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

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

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

an alternative method of moments estimator for $k^{(RP)}$ may be obtained assuming that ρ_t is known. Equating the observed and expected mean squares, we obtain $m\hat{\sigma}^{*2} + \hat{\sigma}_e^2 + \hat{\sigma}_r^2 = MSB$, $\hat{\sigma}_e^2 + \hat{\sigma}_r^2 = MSR$ and $m\hat{\sigma}^{*2} = MSB - MSR$. Then, an alternative method of moments estimator for $k^{(RP)}$ is given by

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

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

2.5.2. *ME* and *SS* Models

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

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

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

$$\mathrm{var}\left(\mathbf{Y}_I^*\right)\mathbf{A} = \frac{1}{n(m-1)}\bigoplus_{i=1}^{n} a_i^2\mathbf{P}_m,$$

and hence

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

Letting $\mathbf{A} = (n-1)^{-1}\left(\mathbf{P}_n \otimes \dfrac{\mathbf{J}_m}{m}\right)$, it follows that

$$\mathrm{var}\left(\mathbf{Y}_I^*\right)\mathbf{A} = \left[m(n-1)\right]^{-1}\left(\left[\bigoplus_{i=1}^{n}\left(a_i^2 + mb^2\right)\right]\mathbf{P}_n\right) \otimes \mathbf{J}_m,$$

and hence

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

When within cluster variances are equal (i.e., $\sigma_i^2 = \sigma_e^2$ for all $i = 1,...,n$), we have

$a_i^2 = \sigma_e^2 + \sigma_r^2$ for all $i = 1,...,n$, and the expected mean squares terms reduce to

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

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

and

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

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

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

for the *ME* model[6], and

$$\begin{aligned}\hat{k}_1^{(SS)}&=\hat{k}^{(ME)}+f\left(1-\hat{k}^{(ME)}\right)\\&=\begin{cases}f & if\ 0\le MSB\le MSR\\ \dfrac{MSB-(1-f)MSR}{MSB} & if\ MSB>MSR\end{cases}\end{aligned}$$

or

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

for the *SS* model.

The empirical shrinkage constants satisfy $0\le\hat{k}^{(ME)}\le\hat{k}_2^{(RP)}\le\hat{k}_2^{(SS)}\le1,$

$0\le\hat{k}_1^{(RP)}\le\hat{k}_2^{(SS)}\le1$ and $0\le\hat{k}^{(ME)}\le\hat{k}_1^{(SS)}\le1.$

[6] 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.

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

3. Details of the simulation study

We conducted a simulation study to compare the *MSE* of different empirical predictors in the context of a two-stage cluster sample from a balanced finite population. The simulation study is 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 for comparative purposes.

3.1. Generation of the finite populations

To encompass a broad number of situations, different compositions for the finite populations are 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 not) of response error is considered at the sampling stage.

We generate each population of units and clusters via the percentiles of some hypothetical distribution. The basic distributions from which we generate the finite populations are normal, uniform, beta, or gamma. These distributions are used only to generate the cluster means; their actual form is not used in the analysis. Although different distributions can be selected for units and clusters, we use the same distribution to generate

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

the unit effects for all clusters in each population. The cluster distribution may or may not agree with that for the units.

For each simulation, the population is composed of N clusters with M units per cluster. We represent each individual cluster parameter by μ_s and their mean by μ. We fix the variance between cluster parameters, σ^2, divide the [0,1] interval into $N+1$ equally spaced intervals and obtain the percentiles corresponding to the upper limit of each interval from the appropriate probability distribution. We redefine the cluster parameters by centering them at μ and re-scaling their values so that the variance matches

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

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

Using σ^2, σ_e^2 and σ_r^2 we define the cluster intra-class and unit intra-class correlation coefficients. We assume that the response error is 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.

Deleted: The response error is assumed

Deleted: SanMartino&al2005v20.doc

Deleted: 18/01/07

Deleted: 09/01/07

Inserted: SanMartino&al2005v20.doc

Deleted: SanMartino&al2005v17.doc

Inserted: 18/01/07

Deleted: 14/02/07

Inserted: 14/02/07

The characteristics of the simulated populations are summarized in Table 1. Fifty-six populations were generated in case 1, corresponding to all combinations of cluster and unit intra-class correlation coefficients. In cases 2 and 3, 336=(6x7x8) populations were generated and 56=(7x8) populations were generated in cases 4 and 5. In total, 840 populations were evaluated.

Insert Table 1 here

3.2. Two-stage cluster sampling

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

For each generated population 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) are considered. This results 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) are considered, resulting in 21 sampling plans for each generated population. For each

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

population and sampling plan, the number of trials is 10000. In total, 8904 different settings for population/sampling plans were simulated.

3.3. Predictors and Empirical Predictors

Once the two-stage samples are obtained, we compute the predictors and empirical predictors as detailed in Section 2.4 and 2.5, respectively. For the *ME* and the *SS* models we assume that the known values of σ^2 and σ_e^2 correspond to the between and within cluster variances, respectively.

Assuming variance components are known, the observed *MSE* is denoted by *SMSE* and under the unknown variance assumption the observed *MSE* is denoted by *EMSE*. The latter correspond to the empirical predictors described in Section 2.6.

4. Simulation Results

To clarify the exposition, we present 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 calculate the relative loss in terms of *EMSE* with respect to *SMSE* that occurs when we replace the theoretical shrinkage constants by

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

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 which presents minimum (*EMS*E) *SMSE*. As sometimes the differences between the (*EMS*E) *SMSE* of two (empirical) predictors is very small, we use the relative percent increase[7] (*RPI*) in (*EMS*E) *SMSE* of each (empirical) predictor relative to the (*EMS*E) *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 criteria for comparing the predictors or the empirical predictors, respectively. Finally, we also identify (empirical) predictors with poor performance (*RPI* > 50%) relative to the best (empirical) predictor.

Initial simulation results indicated that the magnitude of the (*EMS*E) *SMSE* is only slightly affected by changes in the shape of the response distribution. In view of these results, we confine our subsequent analysis to cases 1, 4 and 5 (Table 1). A preliminary analysis also showed that the empirical predictors for the *SS* model with $\hat{k}_1^{(SS)}$ and for the *RP* model with $\hat{k}_1^{(RP)}$ generally have lower *EMSE* than those for which the shrinkage constants are $\hat{k}_2^{(SS)}$ and $\hat{k}_2^{(RP)}$, respectively, so that in subsections 4.2 and 4.3 we restrict the analysis to the former. Due to space limitations, only a few tables and figures are presented

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

[7] The relative percent increase in A relative to B is defined as $\text{RPI} = \frac{\text{A-B}}{\text{B}} \times 100\%$.

for illustration purposes; a more extensive set of tables and figures are available at the web site.

4.1 Performance of predictors

The *RP* model predictor generally presents minimum *SMSE*. Only in a small number of settings, the *ME* or the *SS* model predictors showed minimum *SMSE*, generally for extreme values (0.01 or 0.99) of ρ_s and ρ_t. For these settings, the RPI in *SMSE* of the *RP* model predictor relative to the best predictor is at most 0.03%, which may be justified by the variability introduced by 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 relative percent increase in *SMSE* for the *ME*, the *SS* and the *RP* model predictors with respect to the best predictor (generally that obtained under the *RP* model). The *RP* model predictor is at least equivalent to the best predictor in all the settings. Excluding the *RP* model predictor, the *ME* model predictor is closer to the best predictor, having a lower maximum relative percent increase in *SMSE* than either the *CM* or the *SS* model predictors. As the number of clusters and the cluster sampling fraction increase, the relative percent increase (*RPI*) in *SMSE* also increases.

Insert Table 2 here

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

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 *SS* model predictors have similar *SMSE* to that of the best predictor (generally the *RP* model predictor).

Insert Table 3 here

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

The predictors derived under the *ME* model have poor performance (i.e. *RPI* > 50%) when the cluster intra-class correlation coefficient is small ($\rho_s \leq 0.2$), the unit intra-class correlation coefficient is large ($\rho_t \geq 0.95$) and the unit sampling fraction is large ($f \geq 0.8$) (see Figure A.2 in the web site). Both the *SS* 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 *SS* model predictor and for decreasing f in the case of the *CM* (see Figures A.3 and A.1 in the web site). In contrast, the *RP* model predictor never presents poor performance.

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

4.2 Performance of empirical predictors

4.2.1. Evaluation of the loss due to the use of empirical predictors

To evaluate the loss associated to the use of empirical predictors instead of those where variance components are known, we compute the *RPI* of the *EMSE* with respect to the *SMSE* for each predictor. For the *ME* and the *RP* models, the *EMSE* associated with the empirical predictors always overestimate the *SMSE*, while for the *SS* model empirical predictor, the *EMSE* overestimates the *SMSE* in 85% to 96% of the settings[8]. Table 4 presents some descriptive statistics of the *RPI* of the *EMSE* with respect to the *SMSE* for each predictor, for the simulated populations (see also Figure A.4 in the web site for the corresponding box plots).

Insert Table 4 here

In general, the *SS* model empirical predictor presents a smaller amount of loss than the other two empirical predictors, showing |*RPI*| < 16% in 75% of the settings, followed by

[8] The settings where *EMSE* underestimates *SMSE* generally correspond to situations where $MSB \leq MSR$ for almost all samples. When this happens, we may expect $\bar{Y}^*$ to be a better predictor than $\bar{Y}_i^*$. Additionally, when $MSB \leq MSR$ we generally have that $k^{(SS)} \geq \hat{k}_1^{(SS)}$, so that the theoretical *SS* model predictor puts more weight on $\bar{Y}_i^*$ (and less on $\bar{Y}^*$) than the empirical predictor, producing the superiority of this last one.

Deleted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

the *RP* 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.

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 high unit sampling fractions when $\rho_t \geq 0.8$ and for varying values of ρ_s depending on population and cluster sampling fractions.

4.2.2. Comparison of the *EMSE*

We compute the percentage of settings where each empirical predictor satisfies each of the three criteria previously established: a) minimum *EMS*E, b) "equivalence" to the best empirical predictor (i.e. *RPI* < 15%) and c) poor performance (*RPI* > 50%). It should be noted that these percentages depend on the selection of settings, i.e. the specified population sizes and sampling plans considered in this investigation. Nevertheless, such a summary is one way to provide an overall description of results. All these percentages are calculated considering the combination of all intra-class correlations and unit sampling fractions in the denominator, i.e., 168 (=7x8x3) for the populations with N=10 and M=5 and 392 (=7x8x7) for the others (Tables 5 and 7). In some cases (identified by * in Table 5), the sums of the percentages exceed 100% because the *EMSE* for the *SS* and *RP* model empirical predictors have exactly the same minimum value. In Table 5, the boldfaced

Deleted: 7
Deleted: 8
Deleted: 7
Deleted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

figuresidentify the percentage of settings for the two empirical predictors with the best performance for each type of population and cluster-sampling fraction.

Considering the minimum *EMSE*, the *RP* model empirical predictor does not always have the best performance.

Insert Table 5 here

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

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 empirical one. Table 6 summarizes these results (*RPI* < 15%) for all cluster and unit sampling fractions, but these regions may be extended for specific number and size of clusters and cluster or unit sampling fractions. [9]

Insert Table 6 here

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

Deleted: we also

Deleted: , with red or blue color,

Inserted: we also identify, with red or blue color, the percentage of settings for the two empirical predictors with the best performance for each type of population and cluster-sampling fraction.¶

Deleted: ¶

Deleted: In Table 5 we display the percentage of settings for the two empirical predictors with the best performance for each type of population and cluster-sampling fraction.¶

Deleted: o

Deleted: for those values of cluster and unit correlation coefficients

Formatted

Deleted: SanMartino&al2005v20.doc

Deleted: 18/01/07

Deleted: 09/01/07

Inserted: SanMartino&al2005v20.doc

Deleted: SanMartino&al2005v17.doc

Inserted: 18/01/07

Deleted: 14/02/07

Inserted: 14/02/07

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

It is also important to note that, similar to the case of known variances, both the *CM* and the empirical predictor derived under the *SS* model have poor performance (*RPI* > 50%) in many number of settings (up to 47% and 32%, respectively), followed by the *ME* empirical predictor (up to 7%) (see Table 7).

Insert Table 7 here

Particularly, the empirical predictors derived under the *ME* model show a poor performance when the cluster intra-class correlation coefficient is small ($\rho_s \leq 0.5$), the unit intra-class correlation coefficient is large ($\rho_t \geq 0.8$) and the unit sampling fraction is large ($f \geq 0.6$). For these empirical predictors, the maximum overall *RPI* is 780%. The *SS* 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 *SS* model empirical predictor and for decreasing f in the case of the *CM*. The *RPI* reaches an overall maximum value of 1676% for the *SS* model empirical predictors and 2038% for the *CM*. In contrast, the *RP* 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.

Deleted: 7

Deleted: ¶
Insert Table 7 here¶
¶

Deleted: 8

Deleted: 8

Deleted: not as risky as

Deleted: SanMartino&al2005v20.doc

Deleted: 18/01/07

Deleted: 09/01/07

Inserted: SanMartino&al2005v20.doc

Deleted: SanMartino&al2005v17.doc

Inserted: 18/01/07

Deleted: 14/02/07

Inserted: 14/02/07

5. Discussion

The *RP* 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*.

We also note that, in the absence of response error and when all the units in each sampled cluster are observed, both the *SS* and the *RP* model empirical predictors reproduce the cluster mean, while the *ME* model empirical predictor does not. This highlights the dependence of the *ME* predictor on the assumption of an infinite population.

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

The conclusions obtained in this study are restricted to the case of identical within cluster variances. However another study is in progress considering different within cluster variances and preliminary results are consistent with the present conclusions. Nevertheless, keeping the limitations of simulation studies in mind, our results point in the direction of recommending the *RP* model empirical predictor against its competitors in a variety of settings.

Deleted: s

Deleted: SanMartino&al2005v20.doc
Deleted: 18/01/07
Deleted: 09/01/07
Inserted: SanMartino&al2005v20.doc
Deleted: SanMartino&al2005v17.doc
Inserted: 18/01/07
Deleted: 14/02/07
Inserted: 14/02/07

Acknowledgements

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), USA.

References

Bolfarine, H. and Zacks, S., 1992. *Prediction Theory for Finite Populations*. Springer-Verlag, New York.

Goldberger, A.S., 1962. Best linear unbiased prediction in the generalized linear regression model. *J. Amer. Statist. Assoc.*, **57** 369-375.

Henderson, C.R., 1984. *Applications of Linear Models in Animal Breeding*. Mimeo. (University of Guelph, Guelph, Canada).

McCulloch, C.E. and Searle, S.R., 2001. *Generalized, Linear, and Mixed Effects Models*. John Wiley and Sons, New York.

McLean, R.A., Sanders, W.L. and Stroup, W. W., 1991. A unified approach to mixed linear models. *Amer. Statist.*, **45** 54-64.

Deleted: Cox, D.R. and Solomon, P.J., 2003. *Components of Variance*. Chapman & Hall/CRC, Boca Raton.¶

Deleted: SanMartino&al2005v20.doc

Deleted: 18/01/07

Deleted: 09/01/07

Inserted: SanMartino&al2005v20.doc

Deleted: SanMartino&al2005v17.doc

Inserted: 18/01/07

Deleted: 14/02/07

Inserted: 14/02/07

Robinson, G.K., 1991. That BLUP is a good thing: the estimation of random effects. *Statist. Sci.*, **6** 15-51.

Scott, A. and Smith, T.M.F., 1969. Estimation in multi-stage surveys. *J. Amer. Statist. Assoc.*, **64** 830-840.

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

Searle, S.R. and Fawcett, R.F., 1970. Expected mean squares in variance components models having finite populations. *Biometrics*, **26** 243-254.

Stanek, E.J.III., and Singer, J.M., 2004. Predicting random effects from finite population clustered samples with response error. *J. Amer. Statist. Assoc.* **99** 1119-1130.

Stanek, E.J.III., Singer, J.M. and 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.III., Well, A. and 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.

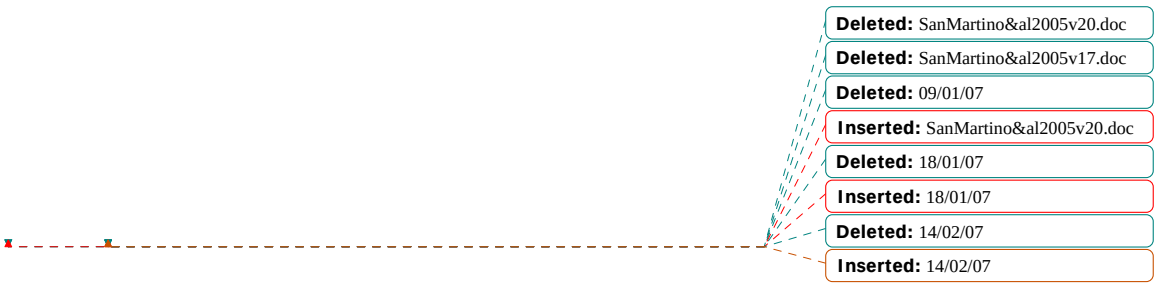

Table captions

Table 1. Characteristics of the simulated populations and sampling plans

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

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

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

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

Table 6. Settings (depending on the intra-class correlation coefficients) where the *CM*, the *ME*, the *SS* and the *RP* empirical predictors have similar performance (*RPI* < 15%) relative to the predictor with minimum *EMSE*

Table 7. Poor performance of each empirical predictor

Deleted: Table 5. Percentage of settings for the two predictors with best performance (minimum *EMSE*)¶

Deleted: SanMartino&al2005v20.doc

Deleted: 18/01/07

Deleted: 09/01/07

Inserted: SanMartino&al2005v20.doc

Deleted: SanMartino&al2005v17.doc

Inserted: 18/01/07

Deleted: 14/02/07

Inserted: 14/02/07

Deleted: 8

Figure captions

Figure 1. Behavior of theoretical shrinkage constants for different values of intra-class correlation coefficients and unit sampling fractions. The plots are organized in such a way that unit intra-class correlation coefficient ρ_t increases from left to right and cluster intra-class correlation coefficient ρ_s increases from top to bottom.