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

Silvina San Martino¹, Julio da Motta Singer², Edward J. Stanek III³

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). Under the general assumptions of the random permutation model and assuming that variances are known and within cluster variances are equal, 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 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.

¹ Facultad de Ciencias Agrarias, Universidad Nacional de Mar del Plata, Mar del Plata, Argentina. silvina@ime.usp.br.
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.
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 never presents a very poor behavior.

*Keywords:* Finite population; Two-stage sampling; Random permutation; Optimal prediction.
1. Introduction

Although authors like Searle, Casella and McCulloch (1992) or Cox and Solomon (2003) mention that finite populations arise quite rarely in applications, there are instances where clustered finite populations occur naturally. This is true 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 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 are generally considered.

The most popular one 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 populations 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 method has had limited application, in part due to its sensitivity to model miss-specification and to 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 extended by Stanek and Singer (2004) to a balanced two-stage sampling with or without response error, considers a design-based probability model induced
by the sampling process. This method does not require restrictive assumptions (such as normality) so that it can be applied to a wide range of practical settings.

Under each model, 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. In certain situations, the predictors obtained under the three models 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 control 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, there may be interest in predicting classroom response. Suppose 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 and the unit intra-class correlation is 0.71). Based on the sample data, we can compute the average response for students in selected classrooms and use the classroom average to estimate the average response for the school. Suppose the school average is 6.75, while for the classroom with teacher \( i \), the sample average is 5.20. 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 the random permutation model predictors. The 11\% observed difference 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 mean squared error (MSE) may be used to select the best predictor.

The mixed effects, Scott and Smith’s and random permutation models rely on different assumptions. Only the last one is based on the sampling scheme and directly links the finite population with a set of random variables without requiring artificial assumptions. 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 (evaluated under the random permutation model assumptions) than those based on the mixed effects and Scott and Smith’s models. From a practical point of view, it may be more realistic to compare the MSE without assuming a specific covariance structure.

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 cluster latent values. We also propose empirical predictors based on estimated variance components. 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 show 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 an individual parameter, is associated with unit $t$ in cluster $s$. We summarize these parameters in the vector $\mathbf{y} = \left( y_1, y'_2, \ldots, y'_N \right)'$ where $y_s = \left( y_{s1}, y_{s2}, \ldots, y_{sM} \right)'$, $s = 1, \ldots, N$.

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 $\frac{M-1}{M} \sigma^2_s = \frac{1}{M} \sum_{t=1}^{M} \left( y_{st} - \mu_s \right)^2$ for $s = 1, \ldots, N$. Also, we let $\sigma^2 = \frac{1}{N} \sum_{s=1}^{N} \sigma^2_s$ 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 $M$ elements in each selected cluster.

### 2.1. The random permutation (RP) model

We consider a probability model induced by the two-stage random sampling process that links the population parameters to an expanded vector of random variables. 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, $i = 1, \ldots, N$ 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...
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 $y_{ij}$ to $Y_{ij}$ we use two indicator random variables: $U_{ij}$, 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 $U_{ij}^{(t)}$, 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

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

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

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

where $\mathbf{Y} = \left( \begin{array}{ccc} \mathbf{Y}_1' & \mathbf{Y}_2' & \cdots & \mathbf{Y}_N' \end{array} \right) \in \mathbb{R}^{NM}$, with $\mathbf{Y}_i = \left( Y_{i1}, Y_{i2}, \ldots, Y_{iM} \right)' \in \mathbb{R}^{M}$,

$$\mathbf{U}^{(t)} = \left( \begin{array}{ccc} \mathbf{U}_1^{(t)} & \mathbf{U}_2^{(t)} & \cdots & \mathbf{U}_M^{(t)} \end{array} \right) \in \mathbb{R}^{M \times NM},$$

with $\mathbf{U}_i^{(t)} = \left( U_{i1}^{(t)}, U_{i2}^{(t)}, \ldots, U_{iM}^{(t)} \right)'$, and

$$\mathbf{U} = \left( \begin{array}{ccc} \mathbf{U}_1 & \mathbf{U}_2 & \cdots & \mathbf{U}_N \end{array} \right) \in \mathbb{R}^{N \times NM},$$

with columns $\mathbf{U}_i = \left( U_{1i}, U_{2i}, \ldots, U_{Mi} \right)'$.

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).
Defining $\beta_s = (\mu_s - \mu)$ as the deviation of the latent value of cluster $s$ from the population mean and $\epsilon_s = (y_s - \mu_s)$ as the deviation of the individual parameter for unit $t$ (in cluster $s$) from the latent value of cluster $s$, we can re-parameterize the vector of fixed values $y$ via the non-stochastic model

$$ y = X\mu + Z\beta + \epsilon $$

(2.1)

where $X = 1_N \otimes 1_M$, $Z = 1_N \otimes 1_M$, $\beta' = (\beta_1, \beta_2, \cdots, \beta_N)$, and $\epsilon$ is defined similarly to $y$.

Here, $1_a$ 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 $X$ and $Z$, it follows that $U1_N = 1_N$, $U^{(i)}1_M = 1_M$, $(U \otimes I_M) \left( \bigoplus_{s=1}^N U^{(s)} \right) X = X$ and $\left( U \otimes I_M \right) \left( \bigoplus_{s=1}^N U^{(s)} \right) Z = U \otimes I_M = ZU$. Then, pre-multiplying both sides of (2.1) by $\left( U \otimes I_M \right) \left( \bigoplus_{s=1}^N U^{(s)} \right)$ and using the above results, we obtain the random permutation mixed effects model

$$ Y = X\mu + ZB + E $$

(2.2)

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

For the random variable $Y$ in (2.2) we have

$$ E_{\mathcal{F}_Y}(Y) = X\mu $$

and

$$ \rho_{\mathcal{F}_Y}(Y) = \Sigma_Y $$

(2.3)
\[ \text{var}_{\xi_1 \xi_2} (Y) = \sigma_x^2 I_{NM} + \sigma_{\varepsilon}^2 (I_N \otimes J_M) - \frac{\sigma^2}{N} J_{NM} \]

where \( \sigma_{\varepsilon}^2 = \frac{\sigma_e^2 \xi}{M} \), \( J_a = I_a I_a' \), and the subscripts \( \xi_1 \) and \( \xi_2 \) 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 \( y \) may not be observed directly, we assume a response error model of the form

\[ \tilde{Y} = Y + W \]

where \( W = (W_s) \in \mathbb{R}^{NM} \) is a vector of independent response errors with \( E(W_s) = 0 \) and \( Var(W_s) = \sigma_{\varepsilon}^2, s = 1, \ldots, N, t = 1, \ldots, M \).

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

\[ Y^* = (U \otimes I_a) \left( \bigoplus_{s=1}^N U(u^{(s)}) \right) \tilde{Y} = Y + W^* \]

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

\[ Y^* = X\mu + ZB + (E + W^*) \, . \tag{2.3} \]

The first two central moments of \( Y^* \) are

\[ E_{\xi_1 \xi_2} (Y^*) = X\mu \]

and

\[ \text{var}_{\xi_1 \xi_2} (Y^*) = (\sigma_x^2 + \sigma_{\varepsilon}^2) I_{NM} + \sigma_{\varepsilon}^2 (I_N \otimes J_M) - \frac{\sigma^2}{N} J_{NM} \, . \]
The subscript $\xi$ denotes expectation with respect to response error and $\sigma^2 = \sum_{i=1}^{N_t} \sum_{j=1}^{M} \frac{\sigma^2_{ij}}{NM}$ denotes the average response error variance.

Note that each realization of the random variable $Y^*$ will generate the same finite population with (two-stage) permuted elements. 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 \textit{realized PSU}. Note also that when there is no response error, the latent value and the cluster mean coincide.

Denoting the sample elements by $Y_i^* = Y_i + W_i^*$, the model for the sample is

$$Y_i^* = X_i \mu + Z_i B + (E_i + W_i^*)$$

where $X_i = I_n \otimes I_m$, $Z_i = I_n \otimes I_m$, $E(Y_i^*) = X_i \mu$ and

$$\text{var}(Y_i^*) = \left( \sigma^2_\xi + \sigma^2_r \right)(I_n \otimes I_m) + \sigma^2_r (I_n \otimes J_m) - \frac{\sigma^2_r}{N}(J_n \otimes J_m).$$

### 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 $y$ is viewed as a realization of a vector of random variables $Y$ such that

$$E(Y) = X \mu \quad \text{and} \quad \text{var}(Y) = \bigoplus_{i=1}^{N}(\sigma^2_{\mu} I_n + \sigma^2_{J_i}). \quad (2.4)$$

Although we use the same notation, neither $\mu$ nor $\sigma^2$ refer to the finite population mean or variance components specified at the beginning of section 2, because the vector $Y$ 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 $i = 1, \ldots, N$,

$$E(\Lambda_i) = \mu \quad \text{and} \quad \text{Var}(\Lambda_i) = \sigma^2. \quad \text{Then,} \quad \sigma^2 \text{ may be interpreted as the variance of the distribution of}$$
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

\[
Y^* = Y + W,
\]

where \( W \sim N\left(0, \sum_{i=1}^{N} \sigma^2_i I_M\right), \) \( i = 1, \ldots, N, \) and \( W \) is independent of \( Y. \) Setting \( \sigma^2_i = \sigma^2, \) \( i = 1, \ldots, N, \) we obtain

\[
E\left(Y^*\right) = X\mu \quad \text{and} \quad \text{var}\left(Y^*\right) = \sum_{i=1}^{N} \left(\sigma^2_i + \sigma^2\right) I_M + \sigma^2 J_M. \quad (2.5)
\]

Denoting the sample elements by \( Y_i^* = Y_i + W_i, \) it follows that \( E\left(Y_i^*\right) = X_i\mu \) and

\[
\text{var}\left(Y_i^*\right) = \sum_{i=1}^{N} \left(\sigma^2_i + \sigma^2\right) I_M + \sigma^2 J_M.
\]

**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 \( PSU \), given by

\[
Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{im})', \ i = 1, \ldots, n, \]

may be modeled by

\[
Y_i = X_i\mu + Z_i B_i + E_i, \quad (2.6)
\]
where \( X_{ji} = Z_{ji} = 1 \) and \( E_{ji} = \left( E_{i1}, E_{i2}, \ldots, E_{im} \right) \). Here, \( Y_{ji} \) is the response of SSU \( j \), \( j = 1, \ldots, m \) in PSU \( i \), \( i = 1, \ldots, n \), \( \mu \) corresponds to the expected response over SSUs and PSUs in the conceptual infinite population, \( B_j \) is a random effect that corresponds to the deviation of the average expected response of SSUs in \( PSU_i \) from \( \mu \), and \( E_{ji} \) 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 \( B_i \overset{\text{iid}}{\sim} N\left(0, \sigma_B^2\right) \) and \( E_{ji} \overset{\text{iid}}{\sim} N\left(0, \sigma^2 \mathbf{I}_m\right) \) are independent, so that \( Y_{ji} \overset{\text{iid}}{\sim} N\left(X_{ji} \mu, \sigma^2 \mathbf{I}_m + \sigma^2 \mathbf{E}_0\right) \). If in (2.6) we include a second additive (response) error term \( W_{ji} \overset{\text{iid}}{\sim} N\left(0, \sigma^2 \mathbf{J}_m\right) \) independent of \( B_i \) and \( E_{ji} \), then
\[
Y^*_j = Y_{ji} + W_{ji} \overset{\text{iid}}{\sim} N\left(X_{ji} \mu, \sigma^2 \mathbf{I}_m + \sigma^2 \mathbf{E}_0 + \sigma^2 \mathbf{J}_m\right).
\]
It follows that
\[
Y_j^* = \left[Y_{j1}^*, Y_{j2}^*, \ldots, Y_{jn}^*\right]' = X_j \mu + Z_j B_j + E_j + W_j \tag{2.7}
\]
where \( B_j = (B_1, B_2, \ldots, B_n)' \), \( E_j = (E_{j1}', E_{j2}', \ldots, E_{jn}')' \), and
\[
W_j = (W_{j1}', W_{j2}', \ldots, W_{jn}')' \text{ which implies that } Y_j^* \overset{\text{iid}}{\sim} N\left(X_j \mu, \sigma^2 \mathbf{I}_m + \sigma^2 \left(\mathbf{J}_m + \mathbf{E}_0\right)\right).
\]
Here again, neither \( \mu \) nor \( \sigma^2 \) and \( \sigma^2_i \) refer to the finite population mean or variance components. Instead, the variance of the random effects, \( \sigma^2_i \), can be interpreted as the variance of the random cluster mean \( \mu + B_j \) 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 \(i \leq n\)), i.e., \(T_i = g'Y\), where \(g' = \frac{1}{M} e'_i \otimes 1_{e'_i}\), and \(e_i\) 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 \(Y_i^* = \left(Y_{i1}^* \ Y_{i2}^* \ \ldots \ Y_{in}^*\right)'\), where \(Y_n^* = \left(Y_{11}^* \ Y_{12}^* \ \ldots \ Y_{mn}^*\right)'\), the predictors of \(T_i\) (for \(i \leq n\)) under the three models (i.e. (2.3), (2.5) or (2.7)) may be obtained as follows:

i) For the mixed effects model:

\[\hat{T}_i = \hat{\mu} + k_i^{(ME)}\left(\overline{Y}_i^* - \hat{\mu}\right),\]

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

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

ii) For the Scott and Smith model:

\[\hat{T}_i = \hat{\mu} + k_i^{(SS)}\left(\overline{Y}_i^* - \hat{\mu}\right),\]

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:

\[\hat{T}_i = \overline{Y}^* + k_i^{(RP)}\left(\overline{Y}_i^* - \overline{Y}^*\right),\]

where \(\overline{Y}^* = \frac{1}{n} \sum_{j=1}^{n} Y_{ij}^*\) and the shrinkage constant is

\[k_i^{(RP)} = \frac{m\sigma^2}{m\sigma^2 + \sigma_i^2 + \sigma_r^2} + \frac{m\sigma^2}{m\sigma^2 + \sigma_r^2 - \left(1 - f\right)\sigma_i^2 + \sigma_r^2}.\]
Under the assumption that the within-cluster variances are identical for all clusters (and equal to $\sigma^2$), the predictors of $T_i$ (for $i \leq n$) under the three models reduce to

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

where

i) For the mixed effects model:

$$k^{(\text{ME})} = \frac{m \sigma^2}{m \sigma^2 + \sigma^2_e + \sigma^2_r}.$$  

ii) For the Scott and Smith model:

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

iii) For the random permutation model:

$$k^{(\text{RP})} = \frac{m \sigma^2}{m \sigma^2 + \sigma^2_e + \sigma^2_r} = \frac{m \sigma^2}{m \sigma^2 + \left(1 - f\right) \sigma^2_e + \sigma^2_r}.$$  

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

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 $f$, $\rho$, and $\rho$ vary, also evaluating the performance of the cluster mean (CM) for which the shrinkage constant is 1. The green solid line identifies the theoretical shrinkage constant for the RP model predictor. 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 $\rho_s$ and $\rho_t$ tend to one. From Figure 1, we also observe that when $f$ tends to 0, $k^{(ME)}$, $k^{(SS)}$ and $k^{(RP)}$ become more similar (and all of them differ from $k^{(CM)}$ (which is equal to one in all cases)).

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 $\rho_t$ tends to zero, $k^{(ME)}$ approaches $k^{(RP)}$ (almost independently of $f$), so that for this situation, we expect that the predictors derived under these two models behave similarly.

Relative to the theoretical results for the RP model predictor, when both $\rho_s$ and $\rho_t$ 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 $\rho_s$ tends to zero and both $\rho_t$ and $f$ tend to one.

### 2.6. 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 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 = \( (n-1)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{m} (\bar{Y}_i - \bar{Y})^2 \) and MSR = \( [n(m-1)]^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} (\bar{Y}_i - \bar{Y})^2 \). These mean squares may be expressed as quadratic forms of the type \( Y''_i A Y'_i \) where \( A = (n-1)^{-1} \left( P_a \otimes \frac{J_m}{m} \right) \) for MSB or \( A = [n(m-1)]^{-1} \left( I_n \otimes P_m \right) \) for MSR, with \( P_a = I_a - a^{-1}J_a \) and \( a \) denoting a positive integer. Since \( A I_{nm} = 0 \) and \( E \left( Y'_i \right) = I_{nm} \mu \), under any of the three models, it follows that

\[
E \left( Y''_i \right) A E \left( Y'_i \right) = 0.
\]

Therefore,

\[
E \left( Y''_i A Y'_i \right) = \text{tr} \left[ \text{var} \left( Y'_i \right) A \right] + E \left( Y''_i \right) A E \left( Y'_i \right) = \text{tr} \left[ \text{var} \left( Y'_i \right) A \right].
\]

Using these results, we may 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 follows.

### 2.6.1. RP model

Under the random permutation model, we have

\[
\text{var} \left( Y'_i \right) = \left( \sigma^2 + \sigma^2_r \right) I_{nm} + \sigma^2 \left( I_n \otimes J_m \right) - \frac{\sigma^2}{N} J_{nm}.
\]

To evaluate the expected value of MSR, we let \( A = [n(m-1)]^{-1} \left( I_n \otimes P_m \right) \), so that

\[
\text{var} \left( Y'_i \right) A = \frac{\left( \sigma^2 + \sigma^2_r \right)}{n(m-1)} \left( I_n \otimes P_m \right),
\]

which implies that

\[
E \left( MSR \right) = \sigma^2 + \sigma^2_r.
\]
To evaluate the expected value of $MSB$, we let $A = (n-1)^{-1} \left( \frac{P_n \otimes J_m}{m} \right)$, so that

$$\text{var} \left( Y_i \right) A = \frac{\sigma_e^2 + \sigma^2}{n-1} \left( \frac{P_n \otimes J_m}{m} \right) + \frac{\sigma^2}{n-1} \frac{P_n \otimes J_m}{m}$$

which implies that

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

Assuming that the response error variance $\sigma_e^2$ is known and equating the observed and expected mean squares, we obtain $\tilde{\sigma}_e^2 + \sigma_i^2 = MSR$ and

$$m\tilde{\sigma}_e^2 = MSB - (1-f)\tilde{\sigma}_e^2 - \sigma_i^2$$
$$= MSB - (1-f)(MSE - \sigma_i^2) - \sigma_i^2.$$
$$= MSB - (1-f)MSE - f\sigma_i^2$$

Consequently, one method of moments estimator for $k^{(RP)}$ is

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

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

$$k^{(RP)} = \frac{m\sigma^2}{m\sigma^2 + \sigma_e^2 + \sigma_i^2} = \frac{m\sigma^2 + f\sigma_i^2}{m\sigma^2 + \sigma_e^2 + \sigma_i^2} = \frac{m\sigma^2 + f\rho_i (\sigma_e^2 + \sigma_i^2)}{m\sigma^2 + \sigma_e^2 + \sigma_i^2},$$

an alternative method of moments estimator for $k^{(RP)}$ may be obtained assuming that

$$\rho_i = \frac{\sigma_e^2}{\sigma_e^2 + \sigma_i^2}$$

is known. Equating the observed and expected mean squares, we obtain

$$m\tilde{\sigma}_e^2 + \tilde{\sigma}_e^2 + \tilde{\sigma}_e^2 = MSB, \quad \tilde{\sigma}_e^2 + \tilde{\sigma}_e^2 = MSR$$
and

$$m\tilde{\sigma}_e^2 = MSB - MSR.$$ 

Then, the alternative method of moments estimator for $k^{(RP)}$ is given by

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

Deleted: 14/02/06
2.6.2. ME and SS Models

Under the ME and SS models, we have

$$\text{var} \left( Y_i^* \right) = \bigotimes_{i=1}^{n} \left( a_i^2 I_m + b_i^2 J_m \right),$$

where $a_i^2 = \sigma^2_i + \sigma^2_r$, and $b_i^2 = \sigma^2_r$. Using $A = [m (m-1)]^{-1} \left( I_n \otimes P_m \right)$, it follows that

$$\text{var} \left( Y_i^* \right) A = \frac{1}{n (m-1)} \bigotimes_{i=1}^{n} a_i^2 P_m,$$

and hence

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

Letting $A = (n-1)^{-1} \left( \begin{array}{c} P_m \otimes J_n \end{array} \right)$, it follows that

$$\text{var} \left( Y_i^* \right) A = \left[ m (n-1) \right]^{-1} \left( \bigotimes_{i=1}^{n} \left( a_i^2 + mb_i^2 \right) \right) \left( P_m \right) \otimes J_m,$$

and hence

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

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

$$E(MSR) = \sigma^2_e + \sigma^2_r$$

and

$$E(MSB) = \sigma^2_e + \sigma^2_r + m \sigma^2_r.$$

In this context, equating observed and expected mean squares, we obtain
\[ \hat{k}^{(ME)} = \frac{m\hat{\sigma}^2}{m\hat{\sigma}^2 + \hat{\sigma}_r^2 + \hat{\sigma}_e^2} = \begin{cases} 0 & \text{if } MSB = 0 \\ \max \left( 0, \frac{MSB - MSR}{MSB} \right) & \text{if } MSB > 0 \end{cases} \]

for the ME model, and

\[ \hat{k}_1^{(SS)} = \hat{k}^{(ME)} + f \left( 1 - \hat{k}^{(ME)} \right) \]

\[ = \begin{cases} f & \text{if } 0 \leq MSB \leq MSR \\ \frac{MSB - (1 - f)MSR}{MSB} & \text{if } MSB > MSR \end{cases} \]

or

\[ \hat{k}_2^{(SS)} = \frac{m\hat{\sigma}^2 + f \left( \hat{\sigma}_r^2 + \hat{\sigma}_e^2 \right)}{m\hat{\sigma}^2 + \hat{\sigma}_r^2 + \hat{\sigma}_e^2} \]

\[ = \begin{cases} f & \text{if } MSB = 0 \\ \max \left( 0, \frac{MSB - (1 - f)MSR}{MSB} \right) & \text{if } MSB > 0 \end{cases} \]

for the SS model.

The empirical shrinkage constants satisfy

\[ 0 \leq \hat{k}^{(ME)} \leq \hat{k}_1^{(RP)} \leq \hat{k}_2^{(SS)} \leq 1, \quad 0 \leq \hat{k}_1^{(RP)} \leq \hat{k}_2^{(SS)} \leq 1 \]

and \( 0 \leq \hat{k}^{(ME)} \leq \hat{k}_1^{(SS)} \leq 1 \).

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

\[ \text{Deleted: SanMartino&al2005v14.doc} \]

\[ \text{Deleted: 14/02/06} \]

\[ ^4 \text{Note that if we assume that the response error variance (} \sigma_r^2 \text{) is known, we would obtain the same estimator for the shrinkage constant, which makes no use of this information.} \]
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, \( \sigma^2 \). 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 the unit effects for all clusters in each population. The cluster distribution may or 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 \( \mu_i \) and their mean by \( \mu \). We fix the variance between cluster parameters, \( \sigma^2 \), 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 \( \mu \) and re-scaling their values so that the variance matches

\[
\sigma^2 = \frac{\sum_{i=1}^{N} (\mu_i - \mu)^2}{N-1}.
\]
Next, we generate unit effects for the \( M \) units within each cluster using percentiles of a specified distribution and force these effects to average zero. The variance of the unit effects is set to be constant for all clusters. The parameters for the cluster units are formed by adding the unit effect to the cluster mean and are represented by \( y_{iu} \). The variance of the unit parameters in cluster

\[
s = \sigma^2 = \frac{1}{M-1} \sum_{i=1}^{M} (y_{iu} - \mu_i)^2.
\]

The common within cluster variance is equal to the average within cluster variance, i.e., \( \sigma^2 = \frac{1}{N} \sum_{i=1}^{N} \sigma^2_{iu} \). Unit effects are re-scaled so that they have zero mean for each cluster and average variance equal to \( \sigma^2 \).

Using \( \sigma^2 \), \( \sigma^2_v \) and \( \sigma^2_e \) we define the cluster intra-class correlation coefficient as

\[
\rho = \frac{\sigma^2}{\sigma^2_v + \sigma^2_e}
\]

and the unit intra-class correlation coefficient as \( r = \frac{\sigma^2}{\sigma^2_v + \sigma^2_e} \). Note that \( \rho = 1 \) 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 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 clusters are combined with...
the population data, and from these data, 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 \( y_{ui} \) 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 \( (F = n / N = 0.2, 0.5 \text{ and } 0.8) \) and three unit sampling fractions \( (f = m / M = 0.4, 0.6 \text{ and } 0.8) \) are considered. This results in nine sampling plans for each generated population. For cases 4 and 5 (Table 1), three cluster sampling fractions \( (F = 0.2, 0.5 \text{ and } 0.8) \) and seven unit sampling fractions \( (f = 0.1, 0.2, 0.4, 0.5, 0.6, 0.8 \text{ and } 0.9) \) are considered, resulting in 21 sampling plans for each generated population. For each 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 \( \sigma^2 \) and \( \sigma^2_c \) 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.
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 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 ($EMSE$) $SMSE$. As sometimes the differences between the ($EMSE$) $SMSE$ of two (empirical) predictors is very small and in simulation studies one should expect some degree of variability, we use the relative percent increase ($RPI$) in ($EMSE$) $SMSE$ of each (empirical) predictor relative 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 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 ($EMSE$) $SMSE$ 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).

---

$^5$ The relative percent increase in A relative to B is defined as $RPI = \frac{A-B}{B} \times 100\%$. 
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 for illustration 
purposes; a more extensive set of tables and figures are available at the web site.

4.1 Performance of predictors

Although the SMSE for each predictor is obtained assuming a different covariance structure 
depending on the model, 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 \( \rho_{\epsilon} \) and \( \rho_{j} \). 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 A.1 in the Appendix 
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.
In Table 2, we summarize general settings (depending on the cluster and unit intra-class correlation coefficients and overall cluster and unit sampling fractions), where the other predictors have similar SMSE to that of the best predictor (generally the RP model predictor).

Insert Table 2 here

From Table 2 we observe that all predictors have similar performance when both intra-class correlation coefficients are high ($\rho_i \geq 0.95$ and $\rho_i \geq 0.5$). The performance of the ME model predictor is more similar to that of the best predictor under a wider range of conditions (i.e. $\rho_i \geq 0.95$ or $\rho_i \leq 0.2$), followed by the SS model predictor when both intra-class correlation coefficients vary (from $\rho_i \geq 0.5$ and $\rho_i \geq 0.8$ to $\rho_i \geq 0.99$ and $\rho_i \geq 0.05$ with one correlation coefficient increasing as the other decreases).

The predictors derived under the ME model have poor performance (i.e. $\text{RPI} > 50\%$) when the cluster intra-class correlation coefficient is small ($\rho_i \leq 0.2$), the unit intra-class correlation coefficient is large ($\rho_i \geq 0.95$) 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.

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 to 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. Table 3 presents some descriptive statistics of the \( RPI \) of the \( EMSE \) with respect to the \( SMSE \) for each predictor, for the simulated populations and Figure 2 shows the corresponding box plots.

Insert Table 3 and Figure 2 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 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 \( \rho_c \geq 0.8 \) and for varying values of \( \rho_s \) 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 EMSE, b) “equivalence” to the best empirical predictor (i.e. RPI < 15%) and c) poor performance (RPI > 50%). 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 A.2 and A.3). In some cases (identified by * in Table A.2), the sums of the percentages exceed 100% because the EMSE for the SS and RP model empirical predictors have exactly the same minimum value.

Considering the minimum EMSE, the RP model empirical predictor does not always have the best performance. In Table 4 we display the percentage of settings for the two empirical predictors with the best performance for each type of population and cluster-sampling fraction.

Insert Table 4 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 $\rho_y$ and $\rho_x$) where each empirical predictor can be considered equivalent (in terms of EMSE) to the best empirical one. Table 5 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.
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 A.2 in the Appendix).

It is also important to note that, similarly to the case of known variances, both the $CM$ and the empirical predictor derived under the $SS$ model have poor performance ($RPI > 50\%$) in a considerable number of settings (5-47% and 10-32%, respectively), followed by the $ME$ empirical predictor (4-7%) (see Table A.3 in the Appendix). Particularly, the empirical predictors derived under the $ME$ model show a poor performance when the cluster intra-class correlation coefficient is small ($\rho_c \leq 0.5$), 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 $CM$ present a poor performance 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 not as risky as its competitors.

5. Discussion
The superiority of the SS model empirical predictor under some of the settings considered in this investigation can be justified for two reasons: 1) the relative loss corresponding to the use of empirical predictors is less for this model than for the RP model and 2) these two predictors have similar performance under a broad range of settings. 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.

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 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 shape of the response distributions under consideration has almost no effect on the values of the MSE.

The conclusions obtained 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.

Finally, keeping the limitations of simulation studies in mind and realizing that additional research may be necessary to obtain more definite conclusions, our results point in the direction of recommending the RP model empirical predictor against its competitors in a variety of settings.

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 Institutes of Health (NIH-PHS-R01-HD36848), USA.

References


APPENDIX – PERFORMANCE OF THE CM AND THE ME, SS AND RP MODEL
(EMPIRICAL) PREDICTORS

1) Known variance components

Insert Table A.1. here

2) Unknown variance components

Insert Table A.2. here

Insert Table A.3. here
Table captions

Table 1. Characteristics of the simulated populations and sampling plans

Table 2. 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 3. Descriptive statistics for the relative percent increase (RPI) in EMSE relative to SMSE for each predictor

Table 4. Percentage of settings for the two predictors with best performance (minimum EMSE)

Table 5. 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 A.1. 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 A.2. Good performance of each empirical predictor

Table A.3. Poor performance of each empirical predictor
**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 $\rho_1$ increases from left to right and cluster intra-class correlation coefficient $\rho_s$ increases from top to bottom.

Figure 2. Relative percent increase in $EMSE$ for the empirical predictors relative to the $SMSE$ for the predictors obtained under the $ME$, the $SS$ or the $RP$ model.