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). When variances are known and within cluster variances are equal, the random permutation model based predictor has smaller mean squared error than the corresponding predictors obtained under 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

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.
of the random permutation model based 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 based empirical predictor is more stable than its competitors because 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 estimation; 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 as in educational, public health or sociological surveys. Children in families, classrooms in schools or physician practices in hospitals are typical examples of such clusters. In such settings, typically, 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., on the averages of the 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. Although 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, these rules have been seldom used, due, in part, to a lack of additional theoretical results and software. 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 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 response 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 true 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.

Under the assumptions that all variances are known and within cluster variances are equal, Stanek and Singer (2004) show that predictors of realized cluster latent value under the random permutation model with response error have smaller MSE than the mixed effects and Scott and Smith’s models. 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 evaluate 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 conclusions, respectively.

2. Predictors of the cluster latent value under different models

We consider a finite population defined by a listing of $M$ 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

\[
y = \left( y_1', \ y_2', \cdots \ y_N' \right)' \quad \text{where} \quad y_s = \left( y_{s1}, \ y_{s2}, \cdots \ y_{sM} \right)', \ s = 1, \ldots, N.
\]

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

\[
\left( \frac{M-1}{M} \right) \sigma_s^2 = \frac{1}{M} \sum_{t=1}^{M} \left( y_{st} - \mu_s \right)^2 \quad \text{for} \ s = 1, \ldots, N.
\]

Also, we let \( \sigma_e^2 = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \) 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 \( M \) elements in each selected cluster.

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

We define 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. In a similar manner, we label the positions in the permutation of units in a cluster by \( j = 1, \ldots, M \). For ease of exposition, we refer to the cluster that will occupy position \( i \) in the permutation of clusters as the \( i^{th} \) primary sampling unit (PSU) and to the unit that will occupy position \( j \) in the permutation of units within a cluster as the \( j^{th} \) secondary sampling unit (SSU).
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

\[
Y_{ij} = \sum_{s=1}^{S} \sum_{t=1}^{T} U_{is} U_{jt}^{(s)} y_{st}.
\]

The finite population \(y\) can be viewed as the realization of the random variable

\[
Y = \left( U \otimes I_M \right) \left( \bigoplus_{s=1}^{S} U_{i}^{(s)} \right) y
\]

where \(Y = \left( \left( Y_{ij} \right) \right) = \left( Y_1' \ Y_2' \ \cdots \ Y_N' \right)' \in \mathbb{R}^{NM}\), with \(Y_i = (Y_{i1} \ Y_{i2} \ \cdots \ Y_{iM})' \in \mathbb{R}^M\),

\(U^{(s)} = \left( U_{1}^{(s)} \ U_{2}^{(s)} \ \cdots \ U_{M}^{(s)} \right) \in \mathbb{R}^{M \times M}\), with \(U_i^{(s)} = \left( U_{1t}^{(s)} \ U_{2t}^{(s)} \ \cdots \ U_{Mt}^{(s)} \right)'\), and

\(U = \left( U_1 \ U_2 \ \cdots \ U_N \right) \in \mathbb{R}^{N \times N}\), with columns \(U_s = \left( U_{1s} \ U_{2s} \ \cdots \ U_{Ns} \right)'\).

Here, \(\otimes\) denotes the Kronecker product and \(\bigoplus_{s=1}^{S} A_s\) denotes a block diagonal matrix with blocks \(A_s\) (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 \(\epsilon_{st} = (y_{st} - \mu_s)\) 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 \( X = I_N \otimes I_M, \ Z = I_N \otimes I_M, \ \beta' = (\beta_1, \beta_2, \ldots, \beta_N), \) 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 \( U1_n = 1_N, \ U^{(s)}1_m = 1_M, \ (U \otimes I_M) \left( \otimes_{x=1}^{N} U^{(s)} \right) X = X \) and

\[
\left( U \otimes I_M \right) \left( \otimes_{x=1}^{N} U^{(s)} \right) Z = U \otimes I_M = ZU.
\]

Then, premultiplying both members of (2.1) by

\[
\left( U \otimes I_M \right) \left( \otimes_{x=1}^{N} U^{(s)} \right)
\]

and using the above results, we obtain the random permutation mixed effects model

\[ Y = X\mu + ZB + E \]

(2.2)

where \( E = \left( U \otimes I_M \right) \left( \otimes_{x=1}^{N} U^{(s)} \right) \epsilon \) and \( B = UB = \left( B_1, B_2, \ldots, B_N \right)' \). Note that because of the random variables \( U \), the terms \( B_i = \sum_{x=1}^{N} U_{ix}B_x \) 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

\[ E_{\hat{\xi}_{y2}}(Y) = X\mu, \]

and

\[ \text{var}_{\hat{\xi}_{y2}}(Y) = \sigma^2 I_{NM} + \sigma^2 \left( I_N \otimes J_M \right) - \frac{\sigma^2}{N} J_{NM} \]

where \( \sigma^2 = \sigma^2 - \frac{\sigma_{\epsilon}^2}{M} \), \( J_a = 1_a 1_a' \), 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 $W = (W_{st}) \in \mathbb{R}^{NM}$ is a vector of independent response errors with $E(W_{st}) = 0$ and $Var(W_{st}) = \sigma^2_{st}$, $s = 1,...,N$, $t = 1,...,M$.

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

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

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

$$Y^* = X\mu + ZB + \left( E + W^* \right).$$

The first two central moments of $Y^*$ are

$$E_{\tilde{Y}^*} \left( Y^* \right) = X\mu$$

and

$$\text{var}_{\tilde{Y}^*} \left( Y^* \right) = \left( \sigma^2 + \sigma^2 \right) I_{NM} + \sigma^2 \left( I_N \otimes J_M \right) - \frac{\sigma^2}{N} J_{NM}.$$ 

The subscript $\tilde{\xi}$ denotes expectation with respect to response error and $\sigma^2 = \sum_{s=1}^N \sum_{t=1}^M \frac{\sigma^2_{st}}{NM}$ 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

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

\textbf{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

$$E(Y) = X \mu \quad \text{and} \quad \text{var}(Y) = \bigoplus_{i=1}^{N} \left(\sigma^2_i I_M + \sigma^2 J_M\right). \quad (2.4)$$

Note that the vector $Y$ in (2.4) is not directly linked to the population units as in the random permutation model. In this context, we have a super-population for the cluster means, i.e., $\mu_1, \ldots, \mu_N$ may be considered as realizations of independent identically distributed random variables $\Lambda_1, \ldots, \Lambda_N$ such that for each $i = 1, \ldots, N$, $E(\Lambda_i) = \mu$ and $\text{Var}(\Lambda_i) = \sigma^2$. Then, $\sigma^2$ 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
\( \mathbf{Y}^* = \mathbf{Y} + \mathbf{W} \),

where \( \mathbf{W} \sim N\left( \mathbf{0}, \sum_{i=1}^{N} \sigma_i^2 \mathbf{1}_M \right) \), \( i = 1, \ldots, N \), \( j = 1, \ldots, M \), and \( \mathbf{W} \) is independent of \( \mathbf{Y} \). Setting

\[ \sigma_n^2 = \sigma_i^2, \quad i = 1, \ldots, N \]

we obtain

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

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

\[
\text{var}\left( \mathbf{Y}_{i}^* \right) = \sum_{i=1}^{n} \left( \left( \sigma_i^2 + \sigma_r^2 \right) \mathbf{1}_m + \sigma_r^2 \mathbf{J}_m \right).
\]

### 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 \( \text{PSU} \, i \), given by

\[ \mathbf{Y}_{i} = (Y_{i1} Y_{i2} \ldots Y_{im})', \quad i = 1, \ldots, n \]

may be modeled by

\[ \mathbf{Y}_{i} = \mathbf{X}_{i}\mu + \mathbf{Z}_{i}B_{i} + \mathbf{E}_{i}, \tag{2.6} \]

where \( \mathbf{X}_{i} = \mathbf{Z}_{i} = \mathbf{1}_m \) and \( \mathbf{E}_{i} = (E_{i1} E_{i2} \ldots E_{im})' \). Here, \( Y_{ij} \) is the response of \( \text{SSU} \, j, \quad j = 1, \ldots, m \) in \( \text{PSU} \, i, \quad i = 1, \ldots, n \), \( \mu \) corresponds to the expected response over \( \text{SSUs} \) and \( \text{PSUs} \) in the conceptual infinite population, \( B_{i} \) is a random effect that corresponds to the deviation of the average expected response of \( \text{SSUs} \) in \( \text{PSU} \, i \) from \( \mu \), and \( E_{ij} \) is a random deviation of the expected response of \( \text{SSU} \, j \) from the average expected response of elements in \( \text{PSU} \, i \). Typically, it is assumed that \( B_{i} \sim N\left( 0, \sigma_B^2 \right) \) and \( E_{ij} \sim N\left( 0, \sigma_E^2 \mathbf{1}_m \right) \) are independent, so that
\[ \mathbf{Y}_{it} \sim N \left( \mathbf{X}_{it} \mu, \sigma_i^2 \mathbf{I}_m + \sigma^2 \mathbf{J}_m \right). \]

If in (2.6) we include a second additive (response) error term

\[ \mathbf{W}_{it} \sim N \left( \mathbf{0}, \sigma_i^2 \mathbf{I}_m \right) \]

independent of \( B_i \) and \( E_{it} \) then

\[ \mathbf{Y}_{it}^* = \mathbf{Y}_{it} + \mathbf{W}_{it} \sim N \left( \mathbf{X}_{it} \mu, \left( \sigma_i^2 + \sigma_r^2 \right) \mathbf{I}_m + \sigma^2 \mathbf{J}_m \right). \]

It follows that

\[ \mathbf{Y}_{it}^* = \left( \mathbf{Y}_{i1}^{*'}, \mathbf{Y}_{i2}^{*'}, \ldots, \mathbf{Y}_{in}^{*'} \right)' = \mathbf{X}_{it} \mu + \mathbf{Z}_t \mathbf{B}_i + \mathbf{E}_t + \mathbf{W}_i \] (2.7)

where \( \mathbf{B}_i = (B_1, B_2, \ldots, B_n)' \), \( \mathbf{E}_t = (\mathbf{E}_{t1}, \mathbf{E}_{t2}, \ldots, \mathbf{E}_{tn})' \), and

\[ \mathbf{W}_i = \left( \mathbf{W}_{i1}', \mathbf{W}_{i2}', \ldots, \mathbf{W}_{in}' \right)' \]

which implies that \( \mathbf{Y}_{it}^* \sim N \left( \mathbf{X}_{it} \mu, \bigoplus_{i=1}^n \left( \sigma_i^2 + \sigma_r^2 \right) \mathbf{I}_m + \sigma^2 \mathbf{J}_m \right) \).

The variance of the random effects, \( \sigma^2 \), can also 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 \( i \leq n \)), i.e., \( T_i = \mathbf{g}' \mathbf{Y}_i \), where \( \mathbf{g}' = \frac{1}{M} \mathbf{e}_i' \otimes \mathbf{I}_{M}' \), 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 \( \mathbf{Y}_i^* = \left( Y_{i1}^*, \ldots, Y_{im}^* \right)' \), where \( \mathbf{Y}_i^* = \left( Y_{i1}^*, \ldots, Y_{im}^* \right)' \), 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:

\[ \hat{T}_i = \hat{\mu} + k_{i}^{(ME)} \left( \bar{Y}_i^* - \hat{\mu} \right), \]
where $Y_i^* = \frac{1}{m} \sum_{j=1}^{m} Y_{ij}^*$, $\hat{\mu} = \frac{\sum_{i=1}^{n} Y_i^* / V_i}{\sum_{i=1}^{n} 1 / V_i}$, $V_i = m\sigma^2 + \sigma_i^2 + \sigma_r^2$, and the shrinkage constant is

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

ii) For the Scott and Smith model:

$$\hat{T}_i = \hat{\mu} + k_i^{(SS)} (\overline{Y}_i^* - \hat{\mu}),$$

where the shrinkage constant is

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

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

iii) For the random permutation model:

$$\hat{T}_i = \overline{Y}^* + k_i^{(RP)} (\overline{Y}_i^* - \overline{Y}^*),$$

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

$$k_i^{(RP)} = \frac{m\sigma^2}{m\sigma^2 + \sigma_i^2 + \sigma_r^2} = \frac{m\sigma^2}{m\sigma^2 + (1 - f) \sigma_i^2 + \sigma_r^2}.$$  

The shrinkage constants satisfy

$$0 \leq k_i^{(ME)} \leq k_i^{(SS)} \leq 1 \quad \text{and} \quad 0 \leq k_i^{(RP)} \leq 1.$$  

Under the assumption that the within-cluster variance is identical for all clusters (and equal to $\sigma_e^2$), the predictors of $T_i$ (for $i \leq n$) under the three models reduce to

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

where:
i) For the mixed effects model:

\[ k^{(ME)} = \frac{m\sigma^2}{m\sigma^2 + \sigma_r^2 + \sigma_e^2}. \]

ii) For the Scott and Smith model:

\[ k^{(SS)} = k^{(ME)} \frac{\sigma_e^2 + \sigma_r^2}{\sigma_r^2} \]

= \frac{m\sigma^2}{m\sigma^2 + \sigma_e^2 + \sigma_r^2}.

iii) For the random permutation model:

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

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

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. We estimate variance components under the limited assumptions of two-stage sampling. These estimators may be derived from ANOVA mean squares, namely

\[ MSB = (n-1) \sum_{j=1}^{n} \sum_{i=1}^{m} (\bar{Y}_i - \bar{Y}^*)^2 \quad \text{and} \quad MSR = [n(m-1)]^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} (Y_{ij}^* - \bar{Y}_i^*)^2. \]

These mean squares may be expressed as quadratic forms of the type \( Y_{i}^{**} A Y_{i}^{*} \) where \( A = (n-1)^{-1} \left( P_n \otimes J_{m} \right) \) 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 \( A 1_{nm} = 0 \) and \( E( Y_{i}^{*} ) = 1_{nm} \mu \), under any of the three models, it follows that \( E( Y_{i}^{**} )A E( Y_{i}^{*} ) = 0 \). Therefore,
Using these results, we may obtain method of moments estimators for the variance components, derive estimators for the shrinkage constants and obtain the corresponding empirical predictors as follows.

### 2.5.1. RP model

Under the random permutation model, we have

\[
\text{var}(Y_i^*) = (\sigma_e^2 + \sigma_r^2)I_{nm} + \sigma_{er}^2 (I_n \otimes J_m) - \frac{\sigma_r^2}{N} J_{nm}.
\]

To evaluate the expected value of MSR, we let

\[
A = [n(m-1)]^{-1} (I_n \otimes P_m),
\]

so that

\[
\text{var}(Y_i^*) A = \frac{\sigma_e^2 + \sigma_r^2}{n(m-1)} (I_n \otimes P_m),
\]

which implies that

\[
E(MSR) = \sigma_e^2 + \sigma_r^2.
\]

To evaluate the expected value of MSB, we let

\[
A = (n-1)^{-1} \left( P_n \otimes J_m \right),
\]

so that

\[
\text{var}(Y_i^*) A = \frac{\sigma_e^2 + \sigma_r^2}{n-1} \left( P_n \otimes J_m \right) + \frac{\sigma_{er}^2}{n-1} P_n \otimes J_m
\]

which implies that

\[
E(MSB) = m\sigma_r^2 + \sigma_e^2 + \sigma_r^2 = m\sigma_e^2 + (1-f)\sigma_e^2 + \sigma_r^2.
\]

Assuming that the response error variance \(\sigma_r^2\) is known and equating the observed and expected mean squares, we obtain \(\hat{\sigma}_e^2 + \hat{\sigma}_r^2 = MSR\) and
\[
m\hat{\sigma}^2 = MSB - (1 - f)\hat{\sigma}_e^2 - \sigma_r^2
\]
\[
= MSB - (1 - f)(MSE - \sigma_r^2) - \sigma_r^2.
\]
\[
= MSB - (1 - f)MSE - f\sigma_r^2.
\]

Consequently, one method of moments estimator for \(k^{(RP)}\) is
\[
\hat{k}_1^{(RP)} = \begin{cases} 
0 & \text{if } MSB = 0 \\
\max\left(0, \frac{MSB - (1 - f)MSR - f\sigma_r^2}{MSB}\right) & \text{if } MSB > 0.
\end{cases}
\]

Re-expressing \(k^{(RP)}\) as
\[
k^{(RP)} = \frac{m\sigma_r^2}{m\sigma_r^2 + \sigma_e^2 + \sigma_r^2} = \frac{m\sigma_r^2 + f\sigma_e^2}{m\sigma_r^2 + \sigma_e^2 + \sigma_r^2} = \frac{m\sigma_r^2 + f\rho\left(\sigma_r^2 + \sigma_e^2\right)}{m\sigma_r^2 + \sigma_e^2 + \sigma_r^2},
\]
an alternative method of moments estimator for \(k^{(RP)}\) may be obtained assuming that \(\rho\) is known.

Equating the observed and expected mean squares, we obtain \(m\hat{\sigma}_e^2 + \hat{\sigma}_r^2 = MSB\), \(\hat{\sigma}_e^2 + \hat{\sigma}_r^2 = MSR\) and \(m\hat{\sigma}_r^2 = MSB - MSR\). Then, the alternative method of moments estimator for \(k^{(RP)}\) is given by
\[
\hat{k}_2^{(RP)} = \begin{cases} 
0 & \text{if } MSB = 0 \\
\max\left(0, \frac{MSB - (1 - f\rho)MSR}{MSB}\right) & \text{if } MSB > 0.
\end{cases}
\]

### 2.5.2. ME and SS Models

Under the ME and SS models, we have
\[
\text{var}\left(Y^*_i\right) = \sum_{i=1}^{n}\left(a_i^2I_m + b^2J_m\right),
\]
where \(a_i^2 = \sigma_i^2 + \sigma_r^2\), and \(b^2 = \sigma_e^2\). Using \(A = [n(m-1)]^{-1}\left(I_n \otimes P_m\right)\), it follows that
\[
\text{var}\left(Y^*_i\right) A = \frac{1}{n(m-1)} \sum_{i=1}^{n} a_i^2 P_m,
\]
and hence

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

Letting \( A = (n-1)^{-1} \left( P_n \otimes \frac{J_m}{m} \right) \), it follows that

\[ \text{var} \left( Y_i^* \right) A = \left[ m(n-1) \right]^{-1} \left( \bigoplus_{i=1}^{n} \left( a_i^2 + mb^2 \right) \right) P_n \otimes J_m, \]

and hence

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

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

\[ E(\text{MSR}) = \sigma_e^2 + \sigma_r^2 \]

and

\[ E(\text{MSB}) = \sigma_e^2 + \sigma_r^2 + m\sigma^2. \]

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

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

for the ME model, and

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

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

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

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

for the SS model.

The empirical shrinkage constants satisfy

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

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

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 \( y_{st} \). The variance of the unit parameters in cluster \( s \) is \( \sigma^2 = \frac{\sum_{t=1}^{M} (y_{st} - \mu_s)^2}{M-1} \). The common within cluster variance is equal to the average within cluster variance, i.e., \( \sigma^2_e = \frac{\sum_{s=1}^{N} \sigma^2_{\sigma_s}}{N} \). Unit effects are re-scaled so that they have zero mean for each cluster and average variance equal to \( \sigma^2_e \).
Using $\sigma^2$, $\sigma^2_e$ and $\sigma^2_r$ we define the cluster intra-class correlation coefficient as

$$\rho_s = \frac{\sigma^2_r}{\sigma^2 + \sigma^2_e}$$

and the unit intra-class correlation coefficient as

$$\rho_t = \frac{\sigma^2_e}{\sigma^2_r + \sigma^2_e}.$$ 

Note that $\rho_t = 1$ corresponds to the case with no response error.

A SAS macro (PSGEN) was developed to generate the finite populations as described in this section and is available from http://www.ime.usp.br/~jmsinger. The user may specify the parameters for the simulation.

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 $y_{st}$ 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 \( f = 0.1, 0.2, 0.4, 0.5, 0.6, 0.8 \text{ and } 0.9 \) 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 \( \sigma_e^2 \) correspond to the between and within cluster variances. To check the simulation results, we compare the observed \( MSE \) to the theoretical mean squared error (TMSE) derived by Stanek and Singer (2004) for the case with known variances.

Under the known variance assumption 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 2.5.

A SAS macro (PRED) was developed to select the two-stage samples, obtain predictors, empirical predictors and observed \( MSE \). It is available from [http://www.ime.usp.br/~jmsinger](http://www.ime.usp.br/~jmsinger).

### 4. Simulation Results

To clarify the exposition, we divide this section into three subsections. In the first subsection we revisit the theoretical results from Stanek and Singer (2004). For equal within cluster
variances, these authors obtain expressions for the $\text{TMSE}$ for the $\text{ME}, \text{SS}$ and $\text{RP}$ model predictors when variances are known and present a small scale simulation study to compare their performance under some settings with varying unit sampling fraction ($f$) and varying cluster and unit intra-class correlation coefficients $\rho_s$ and $\rho_i$. We expand that simulation study by also varying cluster sampling fractions ($F$) and evaluating the performance of the cluster mean ($\text{CM}$) as well. We analyze the behavior of the shrinkage constants and compare the $\text{TMSE}$ of the predictors under each model in order to give a theoretical framework to put the new results, concerning empirical predictors, in perspective.

In the second subsection, we evaluate the relative loss in terms of $\text{EMSE}$ with respect to $\text{TMSE}$ that occurs when we replace the theoretical shrinkage constants by their estimators, i.e., when using the empirical predictors.

Finally, we evaluate the performance of the empirical predictors and determine under what settings each of them has the best performance as well as under what settings they have a bad performance.

A preliminary analysis showed that the empirical predictors for the $\text{SS}$ model with $\hat{k}^{(SS)}_1$ and for the $\text{RP}$ model with $\hat{k}^{(RP)}_1$ generally have lower $\text{EMSE}$ than those for which the shrinkage constants are $\hat{k}^{(SS)}_2$ and $\hat{k}^{(RP)}_2$, 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 set of complete tables and figures can be obtained from http://www.ime.usp.br/~jmsinger.

4.1. Theoretical results revisited
From the expressions derived by Stanek and Singer (2004), predictors obtained under the RP model have minimum TMSE, independently of the number of clusters, number of units or values of cluster or unit intra-class correlation coefficients. As expected, in the simulation study, the RP model empirical predictors have minimum SMSE reproducing the theoretical results. Also, as pointed by Stanek and Singer (2004), the expected TMSE for each predictor does not depend on the shape of the response distribution and it decreases when the cluster or unit sampling fractions increase. The simulation results indicate that values of SMSE are only slightly affected by changes in the shape of the response distribution.

We also use the theoretical results to check the simulation process. With 10,000 trials, the maximum relative percent difference between SMSE and TMSE was 2.7%.

The study of the behavior of the shrinkage constants is a first step to understand the similarities and differences between 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 shrinkage constants as $f$, $\rho_s$ and $\rho_t$ vary. The green solid line identifies the shrinkage constant for the RP model predictor. We consider it as reference to compare the predictors and identify similar as well as different (bad) 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 $f = 0$, $k^{(ME)}$, $k^{(SS)}$ 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 $\rho_t$ 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.

Relatively to the $RP$ model predictor, when both $\rho_t$ and $\rho_s$ tend to zero, the $CM$ should have a bad performance when $f$ approaches zero, while the $SS$ model predictor would have bad performance as $f$ approaches one. Also, we expect the $ME$ model predictor to have a bad performance as $\rho_s$ tends to zero and both $\rho_t$ and $f$ tend to one.

To complete the study of the performance of the different predictors, we also consider a relative comparison of their $TMSE$. For the sampling fractions considered here, Table A.1 in the Appendix and Figures A.1 to A.4 in the web site show the (maximum) relative percent increase in $TMSE^4$ for the $CM$, or the $ME$ and $SS$ model predictors with respect to the $RP$ model predictor.

The $ME$ model predictor is closer to that based on the $RP$ model, having a lower (maximum) relative percent increase in $TMSE$ 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 $TMSE$ also increases.

We also identify general settings for the intra-class correlation coefficients where the $CM$, the $ME$ and the $SS$ model predictors have similar ($RPI < 5\%$) as well as inferior ($RPI > 50\%$) performance relative to the $RP$ model predictor. In Table 2 we indicate the settings (for all cluster and unit sampling fractions) under which the performance of the predictors are similar; some of these settings may be extended depending on the size of the population and the sampling fractions.

---

4 The relative percent increase in A relative to B is defined as $RPI = \frac{A-B}{B} \times 100\%$. 
In Figures A.1 to A.3 (web site), we also show the settings where each predictor has a poor performance for varying cluster and unit sampling fractions.

**Insert Table 2 here**

From Table 2 we observe that all predictors have similar performance when both intra-class correlation coefficients are high (\( \rho_s \geq 0.95 \) and \( \rho_t \geq 0.5 \)). Moreover, all predictors behave similarly for \( \rho_t \geq 0.2 \) when \( \rho_s = 0.99 \). The performance of the ME model predictor is most similar to that of the RP model predictor under a wider range of conditions (i.e. \( \rho_t \geq 0.95 \) 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 when the cluster intra-class correlation coefficient is small (\( \rho_t \leq 0.2 \)), 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 Figures A.2 and A.4 in the web site). Both the SS model predictor and the CM present poor performance 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.1, A.3 and A.4 in the web site).

4.2. **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 $TMSE$ for each predictor. A summary of the results is presented in Table 3.

Insert Table 3 here

In general, the $EMSE$ overestimates the $TMSE$. In particular, the $CM$ presents the lowest maximum $RPI$. Apart from it, the best performance is attained for the empirical predictors under the $SS$ model, with $|RPI| < 20\%$ for about 80-85\% of the settings, while for the $ME$ and $RP$ empirical predictors, this percentage reduces to about 45-70\% and 50-75\%, respectively. The worst performance for all models is attained when unit-sampling fractions are small, especially when both 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 $\rho_i \geq 0.8$ and for varying values of $\rho_s$ depending on population and cluster sampling fractions (see Figure A.5 in the web site). In view of such results, when compared to the known variance case, we expect that:

a) the settings where the $ME$ and $RP$ model predictors are equivalent be maintained or slightly reduced,

b) the number of settings where the $CM$ and the $SS$ model predictor perform well increases,

c) the superiority of $RP$ model predictor is not uniformly preserved.

4.3. Performance of empirical predictors
The *EMSE* of the empirical predictors is not considerably affected when we modify the response distribution. In view of this result, we confine the subsequent analysis to cases 1, 4 and 5 (Table 1).

To evaluate the performance of the empirical predictors we consider three criteria. First, we evaluate the percentage of settings where each empirical predictor attains minimum *EMSE*. As sometimes the differences between the *EMSE* of two (or more) empirical predictors seem to be slight and in simulation studies one should expect some degree of variability, we use an additional criterion (*0 < RPI < 15%*) to identify the percentage of settings where each empirical predictor (including the CM) may be considered “equivalent” to the best empirical predictor (i.e. the one with minimum *EMSE*). Finally, we identify the percentage of settings where each empirical predictor presents poor performance (*RPI > 50%*) relatively to the best one. We also compute the maximum *RPI* (when it exceeds 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 (labeled \( \forall \rho, \rho, f \) in Tables A.2 and A.3). The results are also recomputed with the exclusion of the “extreme” values of the intra-class correlation coefficients (0.01 and 0.99), since such values may inflate or deflate the percentages of settings where the empirical predictors behave similarly (according to the criterion that *0 < RPI < 15%*) or present a poor performance (when *RPI > 50%*). The corresponding columns in Tables A.2 and A.3 are labeled “Excluding \( \rho, \rho = 0.01, 0.99 \)”. As in the general case, we also include the settings where there is no response error (\( \rho = 1 \)). In this context, we use 90 as the denominator for the populations with N=10 and M=5 and 210 for the others. In some cases (identified by *), the percentages exceed 100%, because the *EMSE* for the SS and RP model empirical predictors have exactly the same minimum value.
Considering the minimum \textit{EMSE}, the \textit{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 population and cluster-sampling fraction.

\textbf{Insert Table 4 here}

For populations with \( N = 10 \), the \textit{CM} presents the minimum \textit{EMSE} for small cluster-sampling fractions ( \( F = 0.2 \) ), followed by the \textit{RP} model empirical predictor. As the cluster sampling fraction increases, the \textit{SS} or the \textit{RP} model empirical predictors appear as the two best ones. For populations with \( N = 50 \), the best alternative is generally the \textit{RP} model empirical predictor, followed by the \textit{SS} model empirical predictor.

We also identify certain patterns (depending on \( \rho_s \) and \( \rho_l \) ) where empirical predictors can be considered equivalent (in terms of \textit{EMSE}) to the best one. Table 5 summarizes these results (\textit{RPI} \(< 15\%) for all cluster and unit sampling fractions, but these regions may be extended depending on the number and size of clusters and on the cluster or unit sampling fractions.

\textbf{Insert Table 5 here}

As expected, when we evaluate the loss in terms of \textit{MSE}, the region where the \textit{CM}, the \textit{SS} and the \textit{RP} model empirical predictors are equivalent is expanded relatively to the theoretical framework, while it is reduced for the \textit{ME} model empirical predictor.

Independently of the population characteristics and sampling fractions, the \textit{RP} model empirical predictor is the best or equivalent to the best in a larger number of settings (90 to 100\%) than any competitor.
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%). Particularly, the empirical predictors derived under the ME model show a poor performance when the cluster intra-class correlation coefficient is small ($\rho_z \leq 0.5$), the unit intra-class correlation coefficient is large ($\rho_i \geq 0.8$) 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, the RP model 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 because the loss corresponding to the use of empirical predictors is less for this model than for the RP model. However, 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 is probably justified because it is derived under the assumption of an infinite population.

The optimal performance (in terms of $MSE$) of the $RP$ model predictor derived under the condition of known variances is not entirely reproduced when its empirical version is considered. Nevertheless, 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 does not exhibit a very poor performance. The shape of the response distributions under consideration has almost no effect on the values of the $EMSE$. Also, we restricted our analysis to the case of identical within cluster variances, but another study is in progress considering different within cluster variances; the initial results do not contradict our 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 PREDICTORS UNDER CM, ME, SS AND RP MODELS

1) Known within cluster variances

   Insert Table A.1. here

2) Unknown within cluster variances

   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 \( TMSE \) equivalent to that of the RP model predictor (RPI < 5%)

Table 3. Relative percent increase (RPI) between \( EEMSEMSE \) and \( TMSE \) for each predictor

Table 4. Percentage of settings for the two predictors with best performance

Table 5. Settings where the CM, the ME, the SS and the RP empirical predictors have similar (0 < RPI < 15%) performance relative to the predictor with minimum \( EMSE \)

Table A.1. Maximum relative percent increase (RPI) in expected \( TMSE \) for the CM, the ME and the SS model predictors relative to the RP model predictors for each population

Table A.2. Good performance of each predictor

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