Predicting Random Effects From Finite Population Clustered Samples With Response Error

Edward J. STANEK III and Julio M. SINGER

In many situations there is interest in parameters (e.g., mean) associated with the response distribution of individual clusters in a finite clustered population. We develop predictors of such parameters using a two-stage sampling probability model with response error. The probability model stems directly from finite population sampling without additional assumptions and thus is design-based. The predictors are closely related to best linear unbiased predictors (BLUP) that arise from common mixed-model methods, as well as to model-based predictors obtained via super population approaches for survey sampling. The context assumes clusters of equal size and equal size sampling of units within clusters. Target parameters may correspond to clusters realized in the sample, as well as nonrealized clusters. In either case, the predictors are linear and unbiased, and minimize the expected mean squared error. They correspond to the sum of predictors of responses for realized and nonrealized units in the cluster, accounting directly for the second-stage sampling fraction. In contrast, the BLUP commonly used in mixed models can be interpreted as predicting only the responses of second-stage units not observed for a cluster, not the cluster mean. The development reveals that two-stage sampling does not give rise to a more general variance structure often assumed in superpopulation models, even when variances within clusters are heterogeneous. With response error present, we predict target random variables defined as an expected (or average) response over units in a cluster.

KEY WORDS: Best linear unbiased predictor; Design-based inference; Optimal estimation; Random permutation; Superpopulation; Two-stage sampling.

1. INTRODUCTION

Prediction of random effects in clustered populations is important in many biological and environmental applications. The increased importance has resulted from popularization of mixed-model methods and software, such as SAS PROC MIXED (Littell, Milliken, Stroup, and Wolfinger 1996), for analyzing clustered data, along with the assertion that the predictors of random effects may have useful interpretation. Best linear unbiased predictors (BLUPs) based on mixed-model theory are typically used for realized random effects, that is, random effects associated with realized units (Goldberger 1962; Henderson 1984; McLean, Sanders, and Stroup 1991; Robinson 1991; Stanek, Well, and Ockene 1999; McCulloch and Searle 2001). The sample data are typically considered to have been selected via two-stage sampling from a conceptual infinite population, often vaguely defined but understood to be the limit (as the size becomes infinite) of a finite population of interest.

In many settings, the underlying study population is clustered and finite, and response error is present. Our interest is predicting a value that corresponds to the average of the expected unit response in a cluster (which we call the cluster latent value). Because the clusters and units will be sampled, the parameter of interest may be viewed as a random variable. We also discuss prediction of a related random variable defined by the simple average response of units in a cluster (which we call the cluster mean). We develop predictors of such random variables in situations where each cluster is of the same size, equal size samples of units are selected from each selected cluster, and a single measure of response is made on selected units. For example, defining a physician's practice as a cluster of patients (i.e., units), and a single measure of serum cholesterol on each patient as a response, the average of all the patients' serum cholesterol measures in a practice is the cluster mean, whereas the average of all the patients' expected (over response error) serum cholesterol in the practice is the latent value. When there is no response error, the latent value and the cluster mean coincide.

The approach that we consider closely parallels the prediction-based approach with superpopulation models used in survey sampling (Scott and Smith 1969; Pfeffermann and Nathan 1981; Bolfarine and Zacks 1992; Valliant, Dorfman, and Royall 2000). The prediction approach bases statistical inference on a model for a superpopulation. The finite population is defined as a realization of superpopulation random variables, and the target is prediction of a linear combination of these variables. The actual sample design plays no role in the inference. We present an approach that explicitly develops random variables, similar to a superpopulation, arising from a two-stage cluster sampling design with response error. Because this approach is based on the design, different sources of random variation (e.g., sampling and response error) can be distinguished. This allows us to distinguish the latent value of a cluster from the average response of units in a cluster.

We begin with a brief review of mixed model and superpopulation model results. We next define the basic context and notation for a response error model in a finite population, along with random variables arising from two-stage sampling. We then develop the main results for this model. We conclude by comparing alternative predictors, illustrating them in an example, and discussing more general settings.

1.1 Mixed Models and Superpopulation Sampling Models

In the mixed-model literature, a simple model for the response of the jth second-stage sample unit (SSU), $j = 1$,

Edward J. Stanek III is Professor, Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst, MA 01003. (E-mail: stanek@schoolph.umass.edu). Julio M. Singer is Professor, Department of Statistics, University of São Paulo, São Paulo, Brazil (E-mail: jmsinger@ime.usp.br). Many people contributed to the development of ideas that lead to this manuscript, including Dalton Andrade, Heleno Bolfarine, John Buonaccorsi, Viviana Lencina, and Oscar Loureiro. The authors thank Ira Ockene for providing access to the data from the Seasons study (NIH- R01-HL52745). This work was developed with the support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), and National Institutes of Health (NIH-PHS-R01-HD36848).

Journal of the American Statistical Association
December 2004, Vol. 99, No. 468, Theory and Methods
DOI 10.1198/016214504000001718

$\ldots, m$, in the ith primary sample unit (PSU), $i = 1, \ldots, n$, is given by

$$Y_{ij} = \mu + B_i + E_{ij}, \tag{1}$$

where μ corresponds to the expected response over SSUs and PSUs in a conceptual population, B_i is a random effect corresponding to the deviation from μ of the average expected response of SSUs in the ith PSU, and E_{ij} is a random deviation of the expected response of the jth SSU from the average expected response of the ith PSU. Typically, it is assumed that $B_i \sim \text{iid}\,\text{N}(0, \sigma^2)$ and $E_{ij} \sim \text{iid}\,\text{N}(0, \sigma_i^2)$ are independent. Given (1) and the associated assumptions, the joint distribution of Y_{ij} and B_i may be specified. The corresponding likelihood, when maximized jointly with respect to the fixed and random effects, leads to Henderson's mixed-model equations (Henderson, Kempthorne, Searle, and von Krosigk 1959).

In model (1), assuming known variances, an estimator of the fixed effect μ is the weighted least squares estimator, $\hat{\mu} = \sum_{i=1}^{n} w_i \bar{Y}_i$, where

$$\bar{Y}_i = \frac{1}{m}\sum_{j=1}^{m} Y_{ij}, \qquad w_i = \frac{1/v_i}{\sum_{i=1}^{n} 1/v_i},$$

and

$$v_i = \sigma^2 + \frac{\sigma_i^2}{m}.$$

The solution to the mixed-model equations results in the BLUP of the random effect B_i. Combining the estimator of μ and the predictor of B_i, the predictor of the latent value of the ith realized PSU, $\mu + B_i$, is

$$\hat{\mu} + k_i(\bar{Y}_i - \hat{\mu}), \tag{2}$$

where $k_i = \frac{m\sigma^2}{m\sigma^2 + \sigma_i^2}$.

In practice, the variance parameters are replaced by maximum likelihood or restricted maximum likelihood estimates. McCulloch and Searle (2001) and Robinson (1991) have discussed many other derivations of the BLUP, but none of these derivations account for the impact of a finite population on the predictors.

Predictors of linear combinations of responses of elements of a finite population in a two-stage sampling setting were developed by Scott and Smith (1969) using a superpopulation model, according to which a finite population is viewed as the realization of a set of random variables. The superpopulation is assumed to be nested, with random variables associated with PSUs and SSUs satisfying $E(Y_{ij}) = \mu$ and

$$\text{cov}(Y_{ij}, Y_{kl}) = \begin{cases} \delta^2 + \sigma_i^2 & \text{when } i = k;\ j = l \\ \delta^2 & \text{when } i = k;\ j \neq l \\ 0 & \text{otherwise.} \end{cases}$$

Apart from the difference in notation (using δ^2 in place of σ^2), the first- and second-moment assumptions for the superpopulation match the mixed model assumptions.

A finite population is viewed as the realization of a set of NM random variables indexed by $i = 1, \ldots, N$, $j = 1, \ldots, M$. The parameters of interest are linear combinations of values associated with the elements in the finite population, such as the average response for a PSU. Such parameters can be defined for all PSUs in the finite population. Because only a portion of the PSUs and SSUs are observed, the essential statistical problem is how to predict, in some optimal way, a linear combination of responses of the remaining nonobserved SSUs in the finite population. Predictors are constructed using the joint distribution assumed for the superpopulation (Royall 1976; Bolfarine and Zacks 1992; Valliant et al. 2000). This approach to statistical inference in survey sampling is called model-based, because inference is based on the model assumed for the superpopulation.

Scott and Smith (1969) presented a Bayesian derivation (assuming that the superpopulation is normally distributed), and a distribution-free derivation [based on minimizing the expected mean squared error (MSE) of a linear predictor] that results in the same predictor of the average PSU response given by

$$\hat{P}_i = \frac{m}{M}\bar{Y}_i + \left(\frac{M-m}{M}\right)[\hat{\mu}^* + k_i^*(\bar{Y}_i - \hat{\mu}^*)] \tag{3}$$

with

$$\hat{\mu}^* = \sum_{i=1}^{n} w_i^* \bar{Y}_i \qquad \text{and} \qquad k_i^* = \frac{m\delta^2}{m\delta^2 + \sigma_i^2},$$

where

$$w_i^* = \frac{1/v_i^*}{\sum_{i=1}^{n} 1/v_i^*} \qquad \text{and} \qquad v_i^* = \delta^2 + \frac{\sigma_i^2}{m}$$

if the PSU is in the sample or $\hat{P}_i = \hat{\mu}^*$ if the PSU is not included in the sample. The predictor (3) has an appealing interpretation as the weighted sum of two terms: the sample mean for a PSU in the sample and the predictor of the average of the remaining SSUs for the PSU. The weighting factors are the proportions of the observed and nonobserved SSUs. For PSUs not in the sample, the predictor simplifies to the weighted sample mean.

There is an obvious similarity between (2) and (3). When $\frac{m}{M}$ is small enough so that $1 - \frac{m}{M} \cong 1$, the first term in (3) can be ignored, and the two expressions appear to be identical except for different notation for the variance components. In this setting, one can interpret the BLUP in (2) as a weighted sum of the average response of the observed SSUs and a predictor of the average response of the unobserved SSUs for a PSU, where no weight is given to the mean of the observed SSUs. Such a predictor may be reasonable if the number of SSUs in the PSU is so large that the observed SSUs correspond to a negligible fraction of the total for the PSU.

Scott and Smith's predictor (3) of the average response for a PSU in the sample is a weighted average of the mean response of the sample SSUs and of the predictor of the mean response of the unobserved SSUs for the PSU. This provides a strong intuitive appeal to the prediction-based approach as advocated by Valliant et al. (2000). We borrow the ideas underlying such an approach to develop predictors of the latent value or mean of a PSU directly from a finite-population response error model, using indicator random variables to account for the two-stage sampling.

2. THE TWO–STAGE RANDOM PERMUTATION MODEL FOR A FINITE POPULATION WITH RESPONSE ERROR

The probability model that we consider arises from two-stage sampling of a finite population where a simple response error model is assumed for its units. Sampling is incorporated into the model using indicator random variables that specify a two-stage random permutation of clusters and units in the population, and hence we refer to it as a random permutation model. The model is similar to the superpopulation models discussed by Rao and Bellhouse (1978), who also called their model a random permutation model. We first define a response error model for units in the population and finite-population parameters. We then introduce sampling indicator random variables and define target random variables. We conclude by summarizing the model and specifying the first and second moments.

2.1 The Finite-Population Response Error Model and Parameterizations

Let a finite population be defined by a listing of M units, indexed by $t = 1, \ldots, M$, in each of N clusters, indexed by $s = 1, \ldots, N$, where the kth response for unit t in cluster s is given by

$$Y_{stk} = y_{st} + W_{stk}. \tag{4}$$

This is a response error model where y_{st} is a fixed constant representing the expected response for the unit, and W_{stk} represents response error (with zero expected value). We include the subscript k to help distinguish Y_{stk} in (4) from Y_{ij} in (1), although in our context the subscript could be dropped, because we only consider settings where a single response is made for each unit, that is, $k = 1$. The model can be readily expanded to account for multiple responses per unit.

We denote the response error variance for unit t in cluster s by σ^2_{st}, and define the average response error variance as

$$\sigma^2_r = \sum_{s=1}^{N}\sum_{t=1}^{M}\frac{\sigma^2_{st}}{NM}. \tag{5}$$

The mean and variance of the expected response for units in cluster s are defined as

$$\mu_s = \frac{1}{M}\sum_{t=1}^{M} y_{st}$$

and

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

The parameter μ_s is the latent value for cluster s, and σ^2_s is the survey sampling definition for the variance. We define the average within-cluster variance as

$$\sigma^2_e = \frac{1}{N}\sum_{s=1}^{N}\sigma^2_s. \tag{6}$$

Similarly, we define the population mean and the between-cluster variance as

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

Finally, we define $\beta_s = (\mu_s - \mu)$ as the deviation of the latent value of cluster s from the population mean, and $\varepsilon_{st} = (y_{st} - \mu_s)$ as the deviation of the expected response for unit t in cluster s from the latent value of cluster s. Using these definitions, we represent the expected response for unit t in cluster s $(t = 1, \ldots, M; s = 1, \ldots, N)$ as

$$y_{st} = \mu + \beta_s + \varepsilon_{st}. \tag{7}$$

Model (4) is called a derived model (Hinkelmann and Kempthorne 1994).

Defining

$$\mathbf{y} = (\,\mathbf{y}'_1 \quad \mathbf{y}'_2 \quad \cdots \quad \mathbf{y}'_N\,)',$$

where $\mathbf{y}_s = (y_{s1} \quad y_{s2} \cdots \quad y_{sM})'$, model (7) can be summarized as

$$\mathbf{y} = \mathbf{X}\mu + \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{8}$$

where

$$\mathbf{X} = \mathbf{1}_N \otimes \mathbf{1}_M, \tag{9}$$

$$\mathbf{Z} = \mathbf{I}_N \otimes \mathbf{1}_M, \tag{10}$$

$$\boldsymbol{\beta}' = (\,\beta_1 \quad \beta_2 \quad \cdots \quad \beta_N\,), \tag{11}$$

$\mathbf{1}_a$ is an $a \times 1$ column vector of 1s, $\otimes$ denotes the Kronecker product (Graybill 1983), and $\boldsymbol{\varepsilon}$ is defined similarly to $\mathbf{y}$. Each of the terms in model (8) is a nonstochastic constant. Because $k = 1$, we can define $\mathbf{W} = (W_{stk})$ similarly to $\mathbf{y}$; adding $\mathbf{W}$ to $\mathbf{y}$ summarizes the stochastic finite-population response error model.

2.2 Random Variables and the Two-Stage Random Permutation Model With Response Error

We define the two-stage random permutation model as an ordered list of NM random variables, where both clusters and units in clusters are independently permuted. Assuming that each realization of the two-stage permutation is equally likely, the random variables formally represent two-stage sampling (Cochran 1977).

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 $j = 1, \ldots, M$. Because any unit in any cluster may occupy position ij, we represent the expected response (over response error) for the jth SSU in the ith PSU as the random variable Y_{ij}. For ease of exposition, we refer to the cluster that will occupy position i in the permutation of clusters as the ith PSU, and to the unit that will occupy position j in the permutation of units within a cluster as the jth SSU. PSUs and SSUs are indexed by positions (i and j), whereas clusters and units are indexed by labels (s and t) in the finite population.

To relate y_{st} to Y_{ij}, we use sampling indicator random variables, U_{is}, which take a value of 1 when the ith PSU is

cluster s and a value of 0 otherwise, and indicator random variables $U_{jt}^{(s)}$ that take a value of 1 when the jth SSU in cluster s is unit t and 0 otherwise. As a consequence, the random variable that links the units and clusters to the SSUs and PSUs in a permutation is given by

$$Y_{ij} = \sum_{s=1}^{N} \sum_{t=1}^{M} U_{is} U_{jt}^{(s)} y_{st}. \tag{12}$$

Letting $\mathbf{U}^{(s)} = (\,\mathbf{U}_1^{(s)} \quad \mathbf{U}_2^{(s)} \quad \cdots \quad \mathbf{U}_M^{(s)}\,)$ denote the $M \times M$ matrix with columns

$$\mathbf{U}_t^{(s)} = \left(\, U_{1t}^{(s)} \quad U_{2t}^{(s)} \quad \cdots \quad U_{Mt}^{(s)} \right)',$$

$\mathbf{U} = (\,\mathbf{U}_1 \quad \mathbf{U}_2 \quad \cdots \quad \mathbf{U}_N\,)$ denote the $N \times N$ matrix with columns

$$\mathbf{U}_s = (\,U_{1s} \quad U_{2s} \quad \cdots \quad U_{Ns}\,)',$$

and $\mathbf{Y} = (Y_{ij}) = (\mathbf{Y}_1' \quad \mathbf{Y}_2' \quad \cdots \quad \mathbf{Y}_N')'$ denote an $NM \times 1$ column vector with

$$\mathbf{Y}_i = (\,Y_{i1} \quad Y_{i2} \quad \cdots \quad Y_{iM}\,)',$$

a vector of random variables defining the two-stage permutation of the population is

$$\mathbf{Y} = (\mathbf{U} \otimes \mathbf{I}_M) \left(\bigoplus_{s=1}^{N} \mathbf{U}^{(s)} \right) \mathbf{y}, \tag{13}$$

where $\bigoplus_{s=1}^{N} \mathbf{A}_s$ denotes a block-diagonal matrix with blocks $\mathbf{A}_s$ (Harville 1997). A realization of $\mathbf{Y}$ corresponds to a particular two-stage permutation of the population. The average of the first M random variables in $\mathbf{Y}$ is the latent value for the first PSU.

In a similar manner, we permute the vector of response errors, defining

$$\mathbf{W}^* = (\mathbf{U} \otimes \mathbf{I}_M) \left(\bigoplus_{s=1}^{N} \mathbf{U}^{(s)} \right) \mathbf{W}. \tag{14}$$

Adding (13) to (14) results in

$$\mathbf{Y}^* = \mathbf{Y} + \mathbf{W}^*. \tag{15}$$

This model accounts for the two-stage random permutations and response error. The vector $\mathbf{Y}$ is exchangeable, whereas the vector $\mathbf{Y}^*$ is not, due to the addition of response error. The model for a unit in the jth position in a cluster in the ith position is represented by $Y_{ijk}^* = Y_{ij} + W_{ijk}^*$.

We use the parameterization in (7) and relationship (13) to express the random permutation model as a mixed model. Using the elementary properties of the indicator random variables, it follows that $\mathbf{U}\mathbf{1}_N = \mathbf{1}_N, \mathbf{U}^{(s)}\mathbf{1}_M = \mathbf{1}_M, (\mathbf{U} \otimes \mathbf{I}_M) \times (\bigoplus_{s=1}^{N} \mathbf{U}^{(s)})\mathbf{X} = \mathbf{X}$, and $(\mathbf{U} \otimes \mathbf{I}_M)(\bigoplus_{s=1}^{N} \mathbf{U}^{(s)})\mathbf{Z} = \mathbf{U} \otimes \mathbf{1}_M = \mathbf{Z}\mathbf{U}$. The two-stage permutation of $\mathbf{X}$ is nonstochastic due to the simple structure of $\mathbf{X}$ in (9), whereas the two-stage permutation of $\mathbf{Z}$ can be factored into the product of a nonstochastic matrix times $\mathbf{U}$ due to the simple structure of $\mathbf{Z}$ in (10). Multiplying $\mathbf{U}$ by $\boldsymbol{\beta}$, we define

$$\mathbf{B} = (\,B_1 \quad B_2 \quad \cdots \quad B_N\,)' = \mathbf{U}\boldsymbol{\beta}. \tag{16}$$

The terms $B_i = \sum_{s=1}^{N} U_{is}\beta_s$ for $i = 1, \ldots, N$, are random effects, where B_i represents the deviation of the latent value for PSU i from the population mean. Combining these expressions and defining $\mathbf{E} = (\mathbf{U} \otimes \mathbf{I}_M)(\bigoplus_{s=1}^{N} \mathbf{U}^{(s)})\boldsymbol{\varepsilon}$, we obtain the random permutation mixed model

$$\mathbf{Y}^* = \mathbf{X}\mu + \mathbf{Z}\mathbf{B} + (\mathbf{E} + \mathbf{W}^*). \tag{17}$$

Model (17) differs from the usual mixed model, because it represents all elementary units in the population, as opposed to solely sampled units. The term $(\mathbf{E} + \mathbf{W}^*)$ represents the deviations of response for the SSUs from the PSU latent values.

2.3 Target Random Variables

We assume that there is interest in a linear combination of the random variables $\mathbf{Y}$ of the form

$$T = \mathbf{g}'\mathbf{Y}, \tag{18}$$

where $\mathbf{g}' = (\,\mathbf{g}_1' \quad \mathbf{g}_2' \quad \cdots \quad \mathbf{g}_N'\,)$ with $\mathbf{g}_i = (g_{i1}, \ldots, g_{iM})'$, $i = 1, \ldots, N$, is a vector of known constants. Linear combinations of latent values for PSUs may be defined by taking

$$\mathbf{g}' = \mathbf{b}' \otimes \frac{\mathbf{1}_M'}{M}, \tag{19}$$

where $\mathbf{b} = (b_1, \ldots, b_N)'$ is a vector of known constants. Of principal interest is the linear combination that defines the latent value of PSU i, that is,

$$\mathbf{g}' = \mathbf{e}_i' \otimes \frac{\mathbf{1}_M'}{M}, \tag{20}$$

with $\mathbf{e}_i$ denoting an $N \times 1$ vector with a value of 1 in position i and 0 elsewhere. From (7), (13), (18), and (20), it follows that

$$T = \sum_{s=1}^{N} U_{is}\mu_s. \tag{21}$$

In the context of the superpopulation model described by Scott and Smith (1969), extended by Bolfarine and Zacks (1992) to include response error, the mean of PSU i is defined as

$$T_A = \mathbf{g}'\mathbf{Y}^*, \tag{22}$$

where $T_A = T + \mathbf{g}'\mathbf{W}^*$ and $\mathbf{g}'$ is defined by (20). The random variables (18) and (22) have different interpretations, the difference being the addition of a response error term. A strength of the finite-population sampling-based approach is the ability to distinguish such subtleties in target random variables.

2.4 First and Second Moments of the Random Permutation Mixed Model

Under the two-stage random permutation model (17), using the subscript ξ_1 to denote expectation with respect to permutations of the clusters, the subscript ξ_2 to denote expectation with respect to permutations of units in a cluster, and the subscript ξ_3 to denote expectation with respect to response error, it follows that

$$E_{\xi_1\xi_2\xi_3}(\mathbf{Y}^*) = \mathbf{X}\mu \tag{23}$$

and

$$\operatorname{var}_{\xi_1\xi_2\xi_3}(\mathbf{Y}^*) = (\sigma_r^2 + \sigma_e^2)\mathbf{I}_{NM} + \sigma^{*2}(\mathbf{I}_N \otimes \mathbf{J}_M) - \frac{\sigma^2}{N}\mathbf{J}_{NM}, \tag{24}$$

where

$$\sigma^{*2} = \sigma^2 - \frac{\sigma_e^2}{M}, \tag{25}$$

and $\mathbf{J}_a$ denotes an $(a \times a)$ matrix with all elements equal to 1 (see App. A).

3. PREDICTING THE MEAN AND LATENT VALUE OF A PRIMARY SAMPLE UNIT BASED ON A TWO–STAGE SAMPLE

We assume that m units in each of n clusters are selected by two-stage sampling and that a single measure is obtained on each of the nm SSUs in the sample. The sampling process gives rise directly to the stochastic representation (17) of the population. The random variables in the sample will be realized. Although our principal interest is in predicting the latent value of a realized PSU defined by (18) and (20), we develop unbiased predictors for more general linear combinations of random variables. The predictors are linear functions of the random variables in the sample that minimize the expected value of the mean squared error (MSE). The basic strategy has been given by many authors (e.g., Scott and Smith 1969; Royall 1976; Bolfarine and Zacks 1992; Valliant et al. 2000). We first partition the elements of $\mathbf{Y}^*$ into sampled and remaining portions,

$$\begin{pmatrix} \mathbf{Y}_I^* \\ \hline \mathbf{Y}_{II}^* \end{pmatrix} = \begin{pmatrix} \mathbf{K}_I\mathbf{Y}^* \\ \hline \mathbf{K}_{II}\mathbf{Y}^* \end{pmatrix},$$

premultiplying $\mathbf{Y}^*$ by $\mathbf{K} = (\mathbf{K}_I' \,\vdots\, \mathbf{K}_{II}')'$, where

$$\mathbf{K}_I = \left(\mathbf{I}_n \,\vdots\, \mathbf{0}_{n\times(N-n)}\right) \otimes \left(\mathbf{I}_m \,\vdots\, \mathbf{0}_{m\times(M-m)}\right)$$

and

$$\mathbf{K}_{II} = \begin{pmatrix} \left(\mathbf{I}_n \,\vdots\, \mathbf{0}_{n\times(N-n)}\right) \otimes \left(\mathbf{0}_{(M-m)\times m} \,\vdots\, \mathbf{I}_{M-m}\right) \\ \left(\mathbf{0}_{(N-n)\times n} \,\vdots\, \mathbf{I}_{N-n}\right) \otimes \mathbf{I}_M \end{pmatrix}.$$

We partition $\mathbf{X}$ and $\mathbf{Y}$ in a similar manner, so that

$$\begin{pmatrix} \mathbf{X}_I \\ \hline \mathbf{X}_{II} \end{pmatrix} = \begin{pmatrix} \mathbf{K}_I\mathbf{X} \\ \hline \mathbf{K}_{II}\mathbf{X} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{nm} \\ \hline \mathbf{1}_{NM-nm} \end{pmatrix},$$

$$\begin{pmatrix} \mathbf{Y}_I \\ \hline \mathbf{Y}_{II} \end{pmatrix} = \begin{pmatrix} \mathbf{K}_I\mathbf{Y} \\ \hline \mathbf{K}_{II}\mathbf{Y} \end{pmatrix},$$

and

$$\operatorname{var}_{\xi_1\xi_2\xi_3}\begin{pmatrix} \mathbf{Y}_I^* \\ \hline \mathbf{Y}_{II}^* \end{pmatrix} = \operatorname{var}_{\xi_1\xi_2}\begin{pmatrix} \mathbf{Y}_I \\ \hline \mathbf{Y}_{II} \end{pmatrix} + \sigma_r^2\mathbf{I}_{NM},$$

where

$$\operatorname{var}_{\xi_1\xi_2}\begin{pmatrix} \mathbf{Y}_I \\ \hline \mathbf{Y}_{II} \end{pmatrix} = \begin{pmatrix} \mathbf{V}_I & \mathbf{V}_{I,II} \\ \mathbf{V}_{II,I} & \mathbf{V}_{II} \end{pmatrix},$$

$$\mathbf{V}_I = \sigma_e^2\mathbf{I}_{nm} + \sigma^{*2}(\mathbf{I}_n \otimes \mathbf{J}_m) - \frac{\sigma^2}{N}\mathbf{J}_{nm},$$

$$\begin{aligned}\mathbf{V}_{I,II} &= \mathbf{V}_{II,I}' \\ &= \sigma^{*2}\left(\mathbf{I}_n \otimes \mathbf{J}_{m\times(M-m)} \,\vdots\, \mathbf{0}_{nm\times(N-n)M}\right) \\ &\quad - \frac{\sigma^2}{N}\mathbf{J}_{nm\times(NM-nm)},\end{aligned}$$

and

$$\mathbf{V}_{II} = \sigma_e^2\mathbf{I}_{NM-nm} + \sigma^{*2}\begin{pmatrix} \mathbf{I}_n \otimes \mathbf{J}_{M-m} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N-n} \otimes \mathbf{J}_M \end{pmatrix} - \frac{\sigma^2}{N}\mathbf{J}_{NM-nm}.$$

Finally, we partition $\mathbf{g}'$, resulting in $(\mathbf{g}_I' \,\vdots\, \mathbf{g}_{II}') = (\mathbf{g}'\mathbf{K}_I' \,\vdots\, \mathbf{g}'\mathbf{K}_{II}')$, and partition $\mathbf{g}_{II}'$ corresponding to $\mathbf{K}_{II}'$ as $\mathbf{g}_{II}' = (\mathbf{g}_{1,II}' \,\vdots\, \mathbf{g}_{2,II}')$, where

$$\mathbf{g}_{1,II}' = \mathbf{g}_{II}'\left[\begin{pmatrix} \mathbf{I}_n \\ \hline \mathbf{0}_{(N-n)\times n} \end{pmatrix} \otimes \begin{pmatrix} \mathbf{0}_{m\times(M-m)} \\ \hline \mathbf{I}_{M-m} \end{pmatrix}\right]$$

and

$$\mathbf{g}_{2,II}' = \mathbf{g}_{II}'\left[\begin{pmatrix} \mathbf{0}_{n\times(N-n)} \\ \hline \mathbf{I}_{N-n} \end{pmatrix} \otimes \mathbf{I}_M\right].$$

We assume that the subset $\mathbf{Y}_I^*$ of the elements of $\mathbf{Y}^*$ will be realized, call such a subset "the sample," and express the target random variable T (or T_A) as the sum of two parts. One of these parts depends on the potentially realized sample, and the other depends on the remaining random variables. We require that the predictor of (18) be a linear function of the sampled random variables, that is, that $\hat{T} = (\mathbf{g}_I' + \mathbf{a}')\mathbf{Y}_I^*$ be unbiased, resulting in the constraint that $\mathbf{a}'\mathbf{X}_I - \mathbf{g}_{II}'\mathbf{X}_{II} = 0$, and that it minimize the expected value of the MSE given by $\operatorname{var}_{\xi_1\xi_2\xi_3}(\hat{T} - T)$. Notice that the unbiased constraint requires only that $E_{\xi_1\xi_2\xi_3}(\hat{T}) = E_{\xi_1\xi_2\xi_3}(T)$ (see Robinson 1991). Because cluster sizes and sample sizes within clusters are equal, and a single observation is included in the model for each SSU, the matrices $\mathbf{X}_I$ and $\mathbf{X}_{II}$ are nonstochastic, avoiding the problem noted by Pfeffermann (1984) in Porter's (1973) development of predictors. Letting

$$\hat{T} - T = \left((\mathbf{a}' + \mathbf{g}_I') \,\vdots\, -\mathbf{g}_I' \,\vdots\, -\mathbf{g}_{II}'\right)\begin{pmatrix} \mathbf{Y}_I^* \\ \hline \mathbf{Y}_I \\ \hline \mathbf{Y}_{II} \end{pmatrix}$$

and noting that

$$\operatorname{var}_{\xi_1\xi_2\xi_3}\begin{pmatrix} \mathbf{Y}_I^* \\ \hline \mathbf{Y}_I \\ \hline \mathbf{Y}_{II} \end{pmatrix} = \begin{pmatrix} \mathbf{V}_I^* & \mathbf{V}_I & \mathbf{V}_{I,II} \\ \mathbf{V}_I & \mathbf{V}_I & \mathbf{V}_{I,II} \\ \mathbf{V}_{II,I} & \mathbf{V}_{II,I} & \mathbf{V}_{II} \end{pmatrix},$$

where $\mathbf{V}_I^* = \mathbf{V}_I + \sigma_r^2\mathbf{I}_{nm}$, minimization of $\operatorname{var}_{\xi_1\xi_2\xi_3}(\hat{T} - T)$ subject to the unbiased constraint using Lagrangian multipliers leads to

$$\begin{aligned}\hat{T} &= \mathbf{g}_I'\left[\mathbf{X}_I\hat{\alpha} + \mathbf{V}_I\mathbf{V}_I^{*-1}(\mathbf{Y}_I^* - \mathbf{X}_I\hat{\alpha})\right] \\ &\quad + \mathbf{g}_{II}'\left[\mathbf{X}_{II}\hat{\alpha} + \mathbf{V}_{I,II}'\mathbf{V}_I^{*-1}(\mathbf{Y}_I^* - \mathbf{X}_I\hat{\alpha})\right],\end{aligned} \tag{26}$$

where $\hat{\alpha} = (\mathbf{X}_I'\mathbf{V}_I^{*-1}\mathbf{X}_I)^{-1}\mathbf{X}_I'\mathbf{V}_I^{*-1}\mathbf{Y}_I^*$. Simplifying terms,

$$\hat{\alpha} = \frac{\mathbf{1}_{nm}'}{nm}\mathbf{Y}_I^* = \bar{Y}^*,$$

$$\mathbf{Y}_I^* - \mathbf{X}_I\hat{\alpha} = \mathbf{P}_{nm}\mathbf{Y}_I^*,$$

$$\mathbf{V}_I\mathbf{V}_I^{*-1}\mathbf{P}_{nm} = \left[\rho_t\mathbf{I}_{nm} + (1-\rho_t)k^*\left(\mathbf{I}_n \otimes \frac{\mathbf{J}_m}{m}\right)\right]\mathbf{P}_{nm},$$

and

$$\mathbf{V}'_{I,II}\mathbf{V}_I^{*-1}\mathbf{P}_{nm} = k^*\begin{pmatrix}\mathbf{I}_n \otimes \frac{\mathbf{J}_{(M-m)\times m}}{m}\\ \hline \mathbf{0}_{(N-n)M\times nm}\end{pmatrix}\mathbf{P}_{nm},$$

where

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

and

$$\mathbf{P}_a = \mathbf{I}_a - \frac{\mathbf{J}_a}{a}.$$

Using these expressions, the predictor simplifies to

$$\hat{T} = \mathbf{g}_I'\left[\frac{\mathbf{J}_{nm}}{nm} + \left(\rho_t\mathbf{I}_{nm} + (1-\rho_t)k^*\left(\mathbf{I}_n\otimes\frac{\mathbf{J}_m}{m}\right)\right)\mathbf{P}_{nm}\right]\mathbf{Y}_I^* + \mathbf{g}_{II}'\left[\frac{\mathbf{J}_{(NM-nm)\times nm}}{nm} + k^*\begin{pmatrix}\mathbf{I}_n\otimes\frac{\mathbf{J}_{(M-m)\times m}}{m}\\ \hline \mathbf{0}_{(N-n)M\times nm}\end{pmatrix}\mathbf{P}_{nm}\right]\mathbf{Y}_I^*,$$

which further simplifies to

$$\begin{aligned}\hat{T} = {}& \mathbf{g}_I'\left[\mathbf{1}_{nm}\bar{Y}_I^* + (1-\rho_t)k^*(\mathbf{P}_n\bar{\mathbf{Y}}_I^*\otimes\mathbf{1}_m) + \rho_t\mathbf{P}_{nm}\mathbf{Y}_I^*\right]\\ &+ \mathbf{g}_{1,II}'[\mathbf{1}_{n(M-m)}\bar{Y}_I^* + k^*(\mathbf{P}_n\bar{\mathbf{Y}}_I^*\otimes\mathbf{1}_{M-m})]\\ &+ \mathbf{g}_{2,II}'[\mathbf{1}_{(N-n)M}\bar{Y}_I^*],\end{aligned} \tag{27}$$

where $\bar{\mathbf{Y}}_I^* = (\mathbf{I}_n\otimes\frac{\mathbf{1}_m'}{m})\mathbf{Y}_I^*$ is a vector of PSU sample means. The first term in (27) is the predictor of the linear combination of random variables associated with the realized SSUs for the sample PSUs. The second term is the predictor of the linear combination of SSUs that were not realized for the realized PSUs, whereas the third term is the predictor of the linear combination of the SSUs for PSUs not in the sample. Defining $\mathbf{g}'$ as in (20), the predictor in (27) simplifies to

$$\begin{aligned}\hat{T} = {}& \left(\frac{m}{M}\right)\mathbf{e}_{iI}'\left(\mathbf{1}_n\bar{Y}_I^* + [\rho_t + (1-\rho_t)k^*]\mathbf{P}_n\bar{\mathbf{Y}}_I^*\right)\\ &+ \left(\frac{M-m}{M}\right)\mathbf{e}_{iI}'(\mathbf{1}_n\bar{Y}_I^* + k^*\mathbf{P}_n\bar{\mathbf{Y}}_I^*)\\ &+ (\mathbf{e}_{iII}'\mathbf{1}_{N-n})\bar{Y}_I^*,\end{aligned} \tag{28}$$

where

$$\mathbf{e}_i' = \underset{1\times n}{\left(\mathbf{e}_{iI}'\right.} \;\vdots\; \underset{1\times(N-n)}{\left.\mathbf{e}_{iII}'\right)}.$$

The first two terms in (28) correspond to the predictor of the linear combination of the random variables associated with a sample PSU ($i \le n$); the third term is the predictor of the linear combination of the random variables associated with a PSU that is not in the sample ($i > n$). The term $\rho_t + (1-\rho_t)k^*$ accounts for response error and shrinks the average of the realized values of the random variables associated with the sample SSUs, $\sum_{j=1}^m \frac{1}{m}Y_{ijk}^*$, toward the sample mean when $i \le n$. The term k^* accounts for both response error and between SSU variation and is used to predict the random variables associated with the SSUs not included in the sample by shrinking the PSU average toward the sample mean. A similar result holds for more general linear combinations of random variables associated with PSU latent values; in such cases,

$$\mathbf{e}_i' = \underset{1\times n}{\left(\mathbf{e}_{iI}'\right.} \;\vdots\; \underset{1\times(N-n)}{\left.\mathbf{e}_{iII}'\right)}$$

is replaced by

$$\mathbf{b}' = \underset{1\times n}{\left(\mathbf{b}_I'\right.} \;\vdots\; \underset{1\times(N-n)}{\left.\mathbf{b}_{II}'\right)}$$

in (28).

We compare the predictor of (18) given by (27) to the predictor of (22) given by

$$\begin{aligned}\hat{T}_A = {}& \mathbf{g}_I'\mathbf{Y}_I^* + \mathbf{g}_{1,II}'\left[\mathbf{1}_{n(M-m)}\bar{Y}_I^* + k^*(\mathbf{P}_n\bar{\mathbf{Y}}_I^*\otimes\mathbf{1}_{M-m})\right]\\ &+ \mathbf{g}_{2,II}'[\mathbf{1}_{(N-n)M}\bar{Y}_I^*].\end{aligned} \tag{29}$$

The predictors differ only in the first term. Because the target random variable defined by (18) includes response error, the values associated with the realized SSUs are used directly in the predictor. When $\mathbf{g}'$ is defined by (20), the predictor (29) simplifies to

$$\begin{aligned}\hat{T}_A = {}& \left(\frac{m}{M}\right)\mathbf{e}_{iI}'\bar{\mathbf{Y}}_I^* + \left(\frac{M-m}{M}\right)\mathbf{e}_{iI}'(\mathbf{1}_n\bar{Y}^* + k^*\mathbf{P}_n\bar{\mathbf{Y}}_I^*)\\ &+ [\mathbf{e}_{iII}'\mathbf{1}_{N-n}]\bar{Y}^*.\end{aligned} \tag{30}$$

This predictor is nearly identical to the predictor given by Bolfarine and Zacks (1992), apart from differences in PSU variance parameter definitions.

When there is no response error, $\sigma_r^2 = 0$, and the target random variables given by (18) and (22) are the same. In this case, the predictors given by (28) or (30) simplify to

$$\begin{aligned}\hat{T} = {}& \left(\frac{m}{M}\right)\mathbf{e}_{iI}'\bar{\mathbf{Y}}_I + \left(\frac{M-m}{M}\right)\mathbf{e}_{iI}'(\mathbf{1}_n\bar{Y} + k\mathbf{P}_n\bar{\mathbf{Y}}_I)\\ &+ [\mathbf{e}_{iII}'\mathbf{1}_{N-n}]\bar{Y},\end{aligned} \tag{31}$$

because $\mathbf{e}_{iI}'(\mathbf{1}_n\bar{Y} + \mathbf{P}_n\bar{\mathbf{Y}}_I) = \mathbf{e}_{iI}'\bar{\mathbf{Y}}_I$, where

$$\bar{Y}_i = \frac{\sum_{j=1}^m Y_{ij}}{m},$$

$$\bar{\mathbf{Y}}_I = (\bar{Y}_1 \quad \bar{Y}_2 \quad \cdots \quad \bar{Y}_n)',$$

$$\bar{Y} = \frac{\sum_{i=1}^n\sum_{j=1}^m Y_{ij}}{nm},$$

and

$$k = \frac{m\sigma^{*2}}{m\sigma^{*2}+\sigma_e^2}.$$

This result is nearly identical to the predictor developed by Scott and Smith (1969), with the exception that the shrinkage constant is defined relative to a between-PSU variance component given by (25) as opposed to δ^2. Although the predictors given by (3) and (31) are nearly identical, the model assumptions differ. The predictor (31) is developed assuming a response error model for each unit and a two-stage random permutation of clusters and units in the finite population. These assumptions give rise to different first and second moments than the superpopulation model assumptions that underlie (3). The probability model explicitly represents the two-stage sampling of the finite population.

The expected MSE of the BLUPs can be developed in a straightforward manner (see App. B). When predicting a target random variable defined by (18) with $\mathbf{g}'$ defined by (20), the expression for the expected MSE simplifies, when $i \le n$, to

$$MSE(\hat{T}) = (1 - f\rho_t)\left[\frac{\sigma_e^2}{nm\rho_t} + \left(\frac{n-1}{n}\right)(1-k^*)\sigma^2\right], \quad (32)$$

where $f = \frac{m}{M}$ and, when $i > n$, to

$$MSE(\hat{T}) = \sigma^2\left(1 - \frac{1}{N-n}\right) + \frac{\sigma^{*2}}{n} + \frac{m\sigma^2 + \sigma_e^2 + \sigma_r^2}{nm}. \quad (33)$$

In practice, variance parameters will be unknown and must be estimated. Method-of-moment estimates can be obtained by equating the expected mean squares to the observed mean squares between clusters (MSB) and the observed MSE from an ANOVA table, assuming that σ_r^2 is known. We replace negative variance estimates by 0. The estimate of the within-cluster variance is given by $\hat{\sigma}_e^2 = \max(0, MSE - \sigma_r^2)$. The between-cluster variance is estimated by $\hat{\sigma}^2 = \max(0, \frac{1}{m}[MSB - MSE + f\hat{\sigma}_e^2])$, and we estimate σ^{*2} by $\hat{\sigma}^{*2} = \max(0, \frac{1}{m}[MSB - MSE])$. The MSE of predictors that use estimated variance components will be underestimated by the MSE assuming that the variance components are known, because it will not account for variability in the variance estimates in the predictor. A small-scale simulation study has illustrated that the underestimation may be substantial and presents a complex pattern, motivating the need for additional work in this area.

4. COMPARISON OF PREDICTORS

The predictors developed in Section 3 can be compared directly with predictors developed under mixed model or superpopulation model assumptions. Using common notation and setting $k_r^* = \frac{m\sigma^{*2}+\sigma_e^2}{m\sigma^{*2}+\sigma_e^2+\sigma_r^2}$, when $i \le n$, the predictors of the latent value of PSU i defined by (18) and (20) under different models are summarized in Table 1. Each predictor can be interpreted as the weighted sum of a term predicting the latent values for the SSUs in the sample and a term predicting the latent values for the remaining SSUs. The mixed-model predictor places all of the weight on the second term and is the limit of the predictor proposed by Scott and Smith (1969) under a superpopulation model when the fraction of SSUs observed on a realized PSU is so small that it basically can be neglected. In the balanced two-stage sampling setting considered here, the predictor developed by Scott and Smith (1969) is nearly identical to the predictor based on the random permutation model. The difference between the predictors results from different definitions of variance components and shrinkage constants. The two-stage random permutation model permutes PSUs and hence has a single SSU component of variance representing the average of the SSU within-cluster variances (6), as opposed to the cluster-specific components, σ_i^2. The between-PSU component of variance is defined by (25), as opposed to δ^2. The context described by Scott and Smith (1969) appears to match the context for two-stage sampling, but their model assumptions for the variance components do not correspond to the variance components that arise from permuting clusters and units in a finite population.

When response error is present, the mixed-model and superpopulation model predictors remain the same, but the random permutation model predictor changes. Under the random permutation model with response error, the linear combination of the latent values of the realized SSUs for a selected PSU is predicted using a shrinkage constant, k_r^*, determined by the response error variance.

We simplify the expressions for the predictors in Table 1 under the additional assumptions that the within-cluster variance is identical for all clusters (and equal to σ_e^2), and that the response error variance is identical for all units (and equal to σ_r^2). With these assumptions, $\sigma_i^2 = \sigma_e^2 + \sigma_r^2$. Using a common notation ($\delta^2 = \sigma^2$), we have

$$k_i = k_i^* = \frac{m\sigma^2}{m\sigma^2 + \sigma_e^2 + \sigma_r^2},$$

and each predictor in Table 1 can be expressed as $\hat{T} = \bar{Y}^* + c(\bar{Y}_i^* - \bar{Y}^*)$, where the shrinkage constants c are given in Table 2. When expressed in this form, the differences between the predictors result from the different shrinkage constants. Because values of k_i, k, k^*, and ρ_t are all between 0 and 1, each of the predictors will be closer to the overall sample mean than the simple mean of the realized SSUs for a realized PSU.

Under the random permutation model, the expected MSE of each of the predictors can be expressed as a quadratic function of c given by

$$MSE(\hat{T}) = (1 - f\rho_t)\left[\frac{\sigma_e^2}{nm\rho_t} + \left(\frac{n-1}{n}\right)(1-k^*)\sigma^2\right] + \left(\frac{n-1}{n}\right)\frac{\sigma^{*2}}{k^*}\left(c - [f\rho_t + (1-f\rho_t)k^*]\right)^2. \quad (34)$$

For the random permutation model with response error, the expected MSE simplifies to (32) and (33).

We can use the expression for the expected MSE to compare the expected MSE of different predictors under the two-stage random permutation model. The difference in expected MSE between the mixed-model and random permutation model predictors is given by

$$\mathrm{MSE}(\hat{T}_{\mathrm{MM}}) - \mathrm{MSE}(\hat{T}_{\mathrm{RPR}}) = \left(\frac{n-1}{n}\right)\frac{\sigma^{*2}}{k^*}(c_{\mathrm{MM}} - c_{\mathrm{RPR}})^2,$$

Table 1. Predictors of the Latent Value of PSU i When $i \le n$ in Two-Stage Cluster Sampling

Model	Predictor = Sample SSUs	+	Remaining SSUs
Mixed model	$\hat{p}_i =$		$(\hat{\mu} + k_i(\bar{Y}_i^* - \hat{\mu}))$
Scott and Smith	$\hat{P}_i = f\bar{Y}_i^*$	+	$(1-f)(\hat{\mu}^* + k_i^*(\bar{Y}_i^* - \hat{\mu}^*))$
Random perm.	$\hat{T}_i = f\bar{Y}_i^*$	+	$(1-f)(\bar{Y}^* + k(\bar{Y}_i^* - \bar{Y}^*))$
RP + Resp. err.	$\hat{T}_i = f(\bar{Y}^* + k_r^*(\bar{Y}_i^* - \bar{Y}^*))$	+	$(1-f)(\bar{Y}^* + k^*(\bar{Y}_i^* - \bar{Y}^*))$

Table 2. Values of c for Predictors $\hat{T} = \bar{Y}^* + c(\bar{Y}_i^* - \bar{Y}^*)$ of the Latent Value of PSU i When $i \le n$ in Two-Stage Cluster Sampling With Homogeneous Unit and Response Error Variances

Model	
Mixed model	$c_{\mathrm{MM}} = k_i$
Scott and Smith	$c_{\mathrm{SS}} = f + (1-f)k_i$
Random permutation	$c_{\mathrm{RP}} = f + (1-f)k$
Random permutation with response error	$c_{\mathrm{RPR}} = f\rho_t + (1-f\rho_t)k^*$

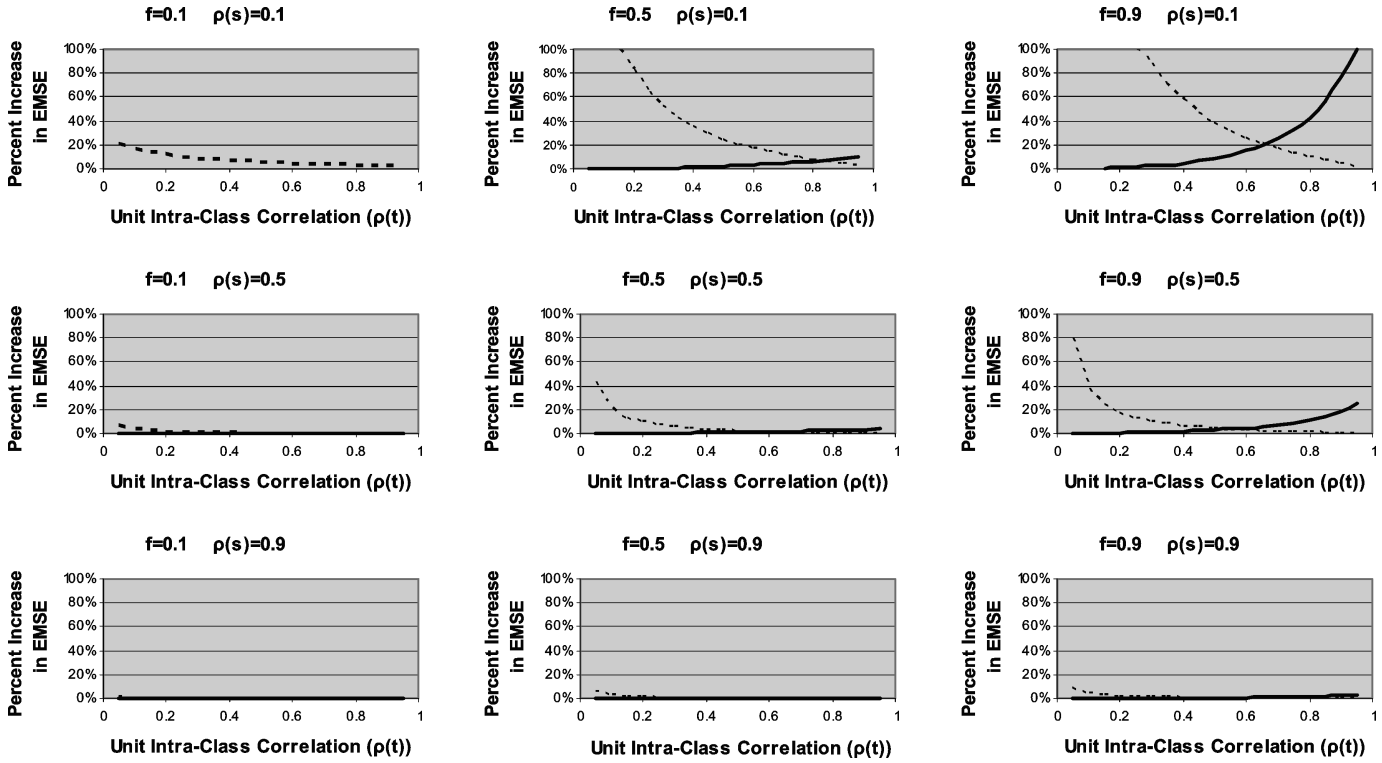

Figure 1. Percent Increase in Expected MSE for Mixed Model (——) and Scott and Smith Model (- - -) Predictors Relative to the Random Permutation Model Predictors of the Latent Value of a Realized Sample PSU by SSU Sampling Fraction (f = .1, .5, .9) and Cluster Intraclass Cluster Correlation ($\rho_s = .1, .5, .9$). Assuming N = 100, n = 30, M = 20, and homogeneous unit and response error variances.

whereas the difference in expected MSE between Scott and Smith's predictor and the random permutation model predictor is given by

$$\mathrm{MSE}(\hat{T}_{\mathrm{SS}}) - \mathrm{MSE}(\hat{T}_{\mathrm{RPR}}) = \left(\frac{n-1}{n}\right)\frac{\sigma^{*2}}{k^*}(c_{\mathrm{SS}} - c_{\mathrm{RPR}})^2.$$

Because each expected MSE is evaluated under the random permutation model assumptions where $\hat{T}_{\mathrm{RPR}}$ is optimal, the differences are always positive. When evaluated under the random permutation model, neither the mixed-model predictor nor Scott and Smith's predictor will have uniformly smaller expected MSE.

Figure 1 illustrates the percent increase in the expected MSE that would occur when using the mixed-model predictor or Scott and Smith's predictor under a two-stage random permutation model. In certain settings, the expected MSE for the mixed-model predictor and Scott and Smith's predictor will exceed that of the random permutation model predictor by more than 100%. The plots in Figure 1 are organized by columns with increasing sampling fractions (f) for a PSU. Rows correspond to increasing cluster intraclass correlations, $\rho_s = \frac{\sigma^2}{\sigma^2+\sigma_e^2}$. The abscissa for the individual plots is the unit intraclass correlation, ρ_t.

Several patterns can be seen in Figure 1. First, Scott and Smith's predictor has high expected MSE for small unit intraclass correlations (ρ_t), especially when the intraclass correlation is modest ($\rho_s < .5$). Mixed-model predictors have high expected MSE when the SSU sampling fraction is large, the unit intraclass correlation is large, and the cluster intraclass correlation is small. When unit intraclass correlations are large ($\rho_t > .75$), Scott and Smith's predictor has smaller expected MSE than the mixed-model predictor. When the SSU sampling fraction is small ($f < .1$), the difference in expected MSE between the mixed-model predictor and the random permutation model predictor is small ($< 5\%$).

5. EXAMPLE

To illustrate our methods, we consider a simple example from a study of seasonal variation in serum cholesterol known as the Seasons study (Merriam, Ockene, Hebert, Rosal, and Matthews 1999). Four or more fasting serum cholesterol levels (with at least one in the summer and in the winter) were collected on a volunteer sample of 20- to 70-year-old members of the Fallon Health Maintenance Organization (HMO) based on 5,000 patient contacts (Ockene et al. 2004) to quantify seasonal patterns in cholesterol. Triplicate 24-hour diet recalls were collected before fasting lipid levels, to control for the impact of diet on cholesterol. A total of 414 subjects had three 24-hour dietary recalls collected in a 6-week period before the first fasting cholesterol measure. The days of collection were selected using a stratified random sample of days in the eligible period, with two weekdays and one weekend day selected. We focus on estimating saturated fat intake for a subject (cluster) corresponding to the first measure of serum cholesterol based on the three 24-hour diet recalls in the 6-week period preceeding the cholesterol measure.

Neither the study subjects (which we consider here as clusters) nor the 24-hour recall days were selected via simple random sampling. Assuming that the 20- to 70-year-old Fallon

HMO membership is 25,000, we estimate the volunteer "participatory" population to be of size $N = 2{,}070$. A stochastic model is needed to make predictions; we assume a two-stage random permutation model to predict average latent saturated fat intake for sampled subjects over the 6-week ($M = 42$ days) period. The model is simple and plausible, but ignores the stratification of weekdays and weekends and assumes that the participating subjects are selected via simple random sampling.

The average saturated fat intake for the $n = 414$ study subjects was 24.8 g/day, with individual subject averages ranging from 2.7 to 92 g/day. The the 95th percentile average saturated fat intake was equal to 50 g/day. A one-way ANOVA model resulted in $MSB = 484$ $(\text{g/day})^2$, and $MSE = 165$ $(\text{g/day})^2$. Unfortunately, the response error variance for saturated fat intake based on a 24-hour recall is not known. Reliability (i.e., ρ_t) is estimated to range from about .4 to .9 (Willett 1990) based on dual 24-hour interviews on the same day. Using these estimates, we assume that the response error variance is $\sigma_r^2 = 25$ $(\text{g/day})^2$ or $\sigma_r^2 = 100$ $(\text{g/day})^2$ when evaluating the predictors.

For illustration, we predict the latent saturated fat intake over the 42-day period for the realized subject whose average 3-day saturated fat intake was 50 g/day. Using method of moments estimates, when $\sigma_r^2 = 25$ $(\text{g/day})^2$, variance components are estimated as $\hat{\sigma}^2 = 110$ $(\text{g/day})^2$ and $\hat{\sigma}_e^2 = 140$ $(\text{g/day})^2$. Using the variance components to estimate c in Table 2, it follows that $\hat{T}_{\text{RPR}} = 41.9$ (g/day). Predictors based on a mixed model [$\hat{T}_{\text{MM}} = 41.5$ (g/day)] or Scott and Smith's model [$\hat{T}_{\text{SS}} = 42.2$ (g/day)] are similar. The expected MSEs of the predictors are also similar, because $f = .071$ and estimates of ρ_s and ρ_t are given by .44 and .85 (see Fig. 1).

Somewhat larger differences occur when we predict the latent saturated fat intake over a 7-day period for the realized subject whose average 3-day saturated fat intake was 50 g/day, assuming that $\sigma_r^2 = 100$ $(\text{g/day})^2$. Now $\hat{\sigma}^2 = 116$ $(\text{g/day})^2$ and $\hat{\sigma}_e^2 = 65$ $(\text{g/day})^2$, so that $\hat{T}_{\text{RPR}} = 42.9$ (g/day). The mixed-model predictor is $\hat{T}_{\text{MM}} = 41.9$ (g/day), and Scott and Smith's predictor is $\hat{T}_{\text{SS}} = 45.3$ (g/day). The expected MSEs of the predictors are once again similar, because $f = .428$ and estimates of ρ_s and ρ_t are given by .64 and .39 (see Fig. 1).

6. DISCUSSION

We have developed predictors of latent values of clusters in a finite population based on a simple probability model that combines a response error model for each unit and a random permutation model for the two-stage finite-population sampling. The model is based on the design, explicitly accounting for the labeled finite population of clusters and units. Others have considered a problem similar to the one presented here, but our approach has substantial differences in terms of formulation of problem and/or the development of a solution.

Konijn (1962) used inclusion probabilities from a two-stage cluster design to define unbiased estimators in a regression model setting. The finite population included a nonstochastic covariate associated with each unit, and in that sense is more general than ours. However, the models considered by Konijn exclude settings where the covariate is equal to 1 for all units in a cluster (which is the setting discussed here). In addition, although Konijn's estimators are unbiased, they have no optimal properties. The predictors developed in Section 3 are based on the design, are unconditionally unbiased, and have minimum MSE.

Pfeffermann and Nathan (1981) extended Scott and Smith's (1969) nested superpopulation model to predict regression coefficients using an approach similar to that we discuss in Section 3. The sample design was not used in defining the model's first and second moments. Similar to the approach of Konijn (1962), a nonstochastic covariate was associated with each unit, and regression parameters differed between clusters. When the covariate is assumed to equal 1 for all units in a cluster, the model considered by Pfeffermann and Nathan is similar to the random permutation model. In this case an important difference between the two models is that Pfeffermann and Nathan's model has zero lack of fit for each unit in a cluster, whereas the random permutation model accounts for lack of fit in the prediction. As noted by Konijn (1962, p. 591), the assumption of perfect fit for such regression models is rarely, if ever, satisfied. The model in Section 2 accounts for lack of fit, and thus the results in Section 3, which differ from Pfeffermann and Nathan's results, have practical relevance.

Other stochastic models have been proposed in this setting. The usual mixed model sidesteps defining a finite population, or clusters and units, while accounting for a hierarchy of random variables via an assumed mean and variance structure. Model-based survey methods do not link the superpopulation to the survey design. Each of the models can be characterized by the first and second moments of a set of random variables (nm random variables in the mixed model and NM random variables in the superpopulation and random permutation models). Predictors are developed in a similar manner in all three models. However, the latent value of a PSU is not readily specified in the superpopulation model when response error is present, and the mixed model does not account for finite population sampling of units. In contrast, the random permutation model with response error can be used to specify the latent value of a PSU while accounting for finite-population sampling.

Appealing to the advantages of the random permutation model, we compare the predictors of the latent value for PSU i (when $i \leq n$) in Table 2 of Section 4. If we view the mixed model and the superpopulation model as earlier attempts to mimic the two-stage sampling, then it is appropriate to evaluate predictors from these models under the random permutation model assumptions. Statistical inference boils down to predicting linear combinations of latent values of units, either observed with error, or unobserved. The BLUP in mixed models (2) is the limit of the predictor proposed by Scott and Smith (1969) under a superpopulation model when the fraction of SSUs observed in a realized PSU is so small that it can basically be neglected. The traditional BLUP places all of the weight on predicting the random variables associated with the unobserved SSUs. Ignoring the proportion of observed SSUs has the disadvantage of increased model sensitivity due to reliance solely on model-based predictors. For example, when there is no response error, if a high fraction of SSUs (say 90%) are observed for a selected PSU, then a large portion of the units composing the PSU are known and need not be predicted. Mixed-model approaches act as if all SSUs must be predicted.

In the balanced two-stage sampling setting considered here, the predictor developed by Scott and Smith (1969) is nearly

identical to the predictor (31) based on the finite-population sampling model. The context described by Scott and Smith appears to match the context for two-stage sampling, but the superpopulation model assumptions for the variance components do not correspond to the two-stage sampling variance components. It is difficult to conceive of a data-generation scheme with clusters and units that would lead to the superpopulation model assumptions. The superpopulation model is in this sense artificial. When response error is present, different target random variables can be specified corresponding to the latent values of a PSU (21) that are not captured in the superpopulation model framework.

The models and methods that we present here are limited to balanced settings with a single measure for each realized SSU. Extensions to settings with different numbers of measures on SSUs, to unbalanced situations, and to settings including other covariables would be desirable. Settings with different numbers of measures on SSUs may be accommodated by taking an average over the measures for a SSU and defining the response error standard deviation by the response error standard error for the SSU. The interpretations of the average response error variance (5) will differ. Note that this extension assumes that the difference in the number of measures per SSU is part of the sampling design.

Extensions to settings where there are different cluster sizes, and possibly different sampling fractions for different clusters, are more complicated. For such extensions when there is no response error, basic random variables that identify both units and positions, that is, $Y_{istj} = U_{is}U_{jt}^{(s)}y_{st}$, are required. Such expanded random variable notation enables the preservation of nesting of subjects in clusters in a random permutation model. Difficulties arise with the expanded random variables due to singularities that have led to representing data for each PSU in the sample and remainder as totals, or averages. Finally, extensions to settings involving covariables on PSUs and/or SSUs are of obvious interest. Such extensions have been developed in a related context with domains, but have not been extended to a two-stage sampling setting (Lencina 2002; Li 2003).

A final caution is warranted when interpreting predictors of random effects. An appealing interpretation of the predictor of a realized random effect is that it predicts the latent value for the realized PSU. Thus if the realized value of U_{is} is equal to 1, then the target parameter given by (21) is μ_s. This interpretation, although seemingly self-evident, may not apply in a two-stage sampling setting. A hint of some problem is apparent in the fact that clusters are not identifiable when representing the two-stage random permutation model as in (13). Dividing the random variables into a sample and remaining portion (as in Sec. 3) will identify clusters that correspond to realized PSUs, because the values of U_{is} will be realized (for $i \le n$) in the sample. When PSUs are of different sizes, some extra care is needed beyond the usual representation as in (13) to verify that the predicted SSUs actually correspond to SSUs for the realized PSU. Aspects of the distinction between a cluster and a PSU are discussed in the context of simple random sampling elsewhere (Lencina 2002; Stanek, Singer, and Lencina 2004). The basic problem can be related to the conventional representation of random variables arising from sampling. In this regard, the work of Godambe (1955) provides a more general framework that may enable a more straightforward strategy for prediction and interpretation.

APPENDIX A: THE VARIANCE UNDER A TWO–STAGE RANDOM PERMUTATION MODEL WITH RESPONSE ERROR

We evaluate the variance of $\mathbf{Y}^*$ using the conditional expansion of the variance, that is, $\operatorname{var}_{\xi_1\xi_2\xi_3}(\mathbf{Y}^*) = E_{\xi_1\xi_2}[\operatorname{var}_{\xi_3|\xi_1\xi_2}(\mathbf{Y}^*)] + \operatorname{var}_{\xi_1\xi_2}[E_{\xi_3|\xi_1\xi_2}(\mathbf{Y}^*)]$. Given ξ_1 and ξ_2 and using (14) and (15), because $\mathbf{Y}$ is nonstochastic, it follows that

$$\operatorname{var}_{\xi_3|\xi_1\xi_2}(\mathbf{Y}^*) = \left[(\mathbf{U}\otimes\mathbf{I}_M)\left(\bigoplus_{s=1}^{N}\mathbf{U}^{(s)}\right)\right] \times \operatorname{var}_{\xi_3|\xi_1\xi_2}(\mathbf{W})\left[(\mathbf{U}\otimes\mathbf{I}_M)\left(\bigoplus_{s=1}^{N}\mathbf{U}^{(s)}\right)\right]'.$$

Now,

$$\operatorname{var}_{\xi_3|\xi_1\xi_2}(\mathbf{W}) = \mathbf{D}_{\mathbf{r}^2} = \bigoplus_{s=1}^{N}\mathbf{D}_{\mathbf{r}_s^2},$$

where $\mathbf{r}^2 = (\mathbf{r}_1^{2\prime} \vdots \mathbf{r}_2^{2\prime} \vdots \cdots \vdots \mathbf{r}_N^{2\prime})'$, $\mathbf{r}_s^2 = (\sigma_{s1}^2 \quad \sigma_{s2}^2 \quad \cdots \quad \sigma_{sM}^2)'$, and $\mathbf{D}_{\mathbf{a}}$ represents a diagonal matrix with diagonal elements equal to the elements of $\mathbf{a}$. Thus

$$\operatorname{var}_{\xi_3|\xi_1\xi_2}(\mathbf{Y}^*) = \left[(\mathbf{U}\otimes\mathbf{I}_M)\left(\bigoplus_{s=1}^{N}\mathbf{U}^{(s)}\right)\right]\mathbf{D}_{\mathbf{r}^2}\left[(\mathbf{U}\otimes\mathbf{I}_M)\left(\bigoplus_{s=1}^{N}\mathbf{U}^{(s)}\right)\right]'.$$

Because $E_{\xi_3|\xi_1\xi_2}(\mathbf{Y}^*) = \mathbf{Y}$,

$$\begin{aligned}\operatorname{var}_{\xi_1\xi_2\xi_3}(\mathbf{Y}^*) &= E_{\xi_1\xi_2}\left[\left[(\mathbf{U}\otimes\mathbf{I}_M)\left(\bigoplus_{s=1}^{N}\mathbf{U}^{(s)}\right)\right]\mathbf{D}_{\mathbf{r}^2}\left[(\mathbf{U}\otimes\mathbf{I}_M)\left(\bigoplus_{s=1}^{N}\mathbf{U}^{(s)}\right)\right]'\right] \\ &\quad + \operatorname{var}_{\xi_1\xi_2}[\mathbf{Y}], \qquad (A.1)\end{aligned}$$

with $\mathbf{Y}$ defined by (13).

Using the conditional expansion, we express the first term in as

$$E_{\xi_1}\left[(\mathbf{U}\otimes\mathbf{I}_M)E_{\xi_2|\xi_1}\left[\left(\bigoplus_{s=1}^{N}\mathbf{U}^{(s)}\right)\mathbf{D}_{\mathbf{r}^2}\left(\bigoplus_{s=1}^{N}\mathbf{U}^{(s)}\right)'\right](\mathbf{U}\otimes\mathbf{I}_M)'\right].$$

Then

$$E_{\xi_2|\xi_1}\left[\left(\bigoplus_{s=1}^{N}\mathbf{U}^{(s)}\right)\mathbf{D}_{\mathbf{r}^2}\left(\bigoplus_{s=1}^{N}\mathbf{U}^{(s)}\right)'\right] = \bigoplus_{s=1}^{N}E_{\xi_2|\xi_1}[\mathbf{U}^{(s)}\mathbf{D}_{\mathbf{r}_s^2}\mathbf{U}^{(s)\prime}].$$

The matrix in parentheses is a diagonal matrix with diagonal elements interchanged for all positions on the diagonal. As a result, defining $\bar{\sigma}_s^2 = \sum_{t=1}^{M}\frac{\sigma_{st}^2}{M}$, the expectation is given by $E_{\xi_2|\xi_1}[\mathbf{U}^{(s)}\mathbf{D}_{\mathbf{r}_s^2}\mathbf{U}^{(s)\prime}] = \bar{\sigma}_s^2\mathbf{I}_M$. Thus

$$\begin{aligned}&E_{\xi_1\xi_2}\left[\left[(\mathbf{U}\otimes\mathbf{I}_M)\left(\bigoplus_{s=1}^{N}\mathbf{U}^{(s)}\right)\right]\mathbf{D}_{\mathbf{r}^2}\left[(\mathbf{U}\otimes\mathbf{I}_M)\left(\bigoplus_{s=1}^{N}\mathbf{U}^{(s)}\right)\right]'\right] \\ &= E_{\xi_1}\left[(\mathbf{U}\otimes\mathbf{I}_M)\left(\bigoplus_{s=1}^{N}\bar{\sigma}_s^2\mathbf{I}_M\right)(\mathbf{U}\otimes\mathbf{I}_M)'\right] \\ &= E_{\xi_1}\left(\mathbf{U}\mathbf{D}_{\bar{\sigma}_s^2}\mathbf{U}'\right)\otimes\mathbf{I}_M,\end{aligned}$$

where $\mathbf{D}_{\bar{\sigma}_s^2}$ is an $N\times N$ diagonal matrix with elements $\bar{\sigma}_s^2$ on the diagonal. In a similar manner, we find that $E_{\xi_1}(\mathbf{U}\mathbf{D}_{\bar{\sigma}_s^2}\mathbf{U}')\otimes\mathbf{I}_M = \sigma_r^2\mathbf{I}_{NM}$,

where $\sigma_r^2 = \sum_{s=1}^N \frac{\bar{\sigma}_s^2}{N}$. As a result,

$$E_{\xi_1\xi_2}\left[\left[(\mathbf{U}\otimes\mathbf{I}_M)\left(\bigoplus_{s=1}^N \mathbf{U}^{(s)}\right)\right]\mathbf{D}_{\mathbf{r}^2} \times\left[(\mathbf{U}\otimes\mathbf{I}_M)\left(\bigoplus_{s=1}^N \mathbf{U}^{(s)}\right)\right]'\right] = \sigma_r^2\mathbf{I}_{NM}. \quad \text{(A.2)}$$

Next we evaluate $\text{var}_{\xi_1\xi_2}(\mathbf{Y}) = \text{var}_{\xi_1}[E_{\xi_2|\xi_1}(\mathbf{Y})] + E_{\xi_1}[\text{var}_{\xi_2|\xi_1}(\mathbf{Y})]$. The first term in this expansion is given by $\text{var}_{\xi_1}[E_{\xi_2|\xi_1}(\mathbf{Y})] = \text{var}_{\xi_1}(\mathbf{X}\mu + \mathbf{ZB})$. Now $\text{var}_{\xi_1}[E_{\xi_2|\xi_1}(\mathbf{Y})] = \text{var}_{\xi_1}(\mathbf{U}\boldsymbol{\beta})\otimes\mathbf{J}_M$, where $\text{var}_{\xi_1}(\mathbf{U}\boldsymbol{\beta})$ is equal to the variance of a random permutation of cluster latent values in the population. As a result, $\text{var}_{\xi_1}[E_{\xi_2|\xi_1}(\mathbf{Y})] = \sigma^2(\mathbf{P}_N\otimes\mathbf{J}_M)$, where $\mathbf{P}_a = \mathbf{I}_a - \frac{\mathbf{J}_a}{a}$. We evaluate the second term in the conditional expansion of the variance in a similar manner. For the cluster in position i in a given permutation of clusters, the variance of the SSUs is

$$\text{var}_{\xi_2|\xi_1}\begin{pmatrix}\mathbf{Y}_i\\ M\times 1\end{pmatrix} = s_i^2\mathbf{P}_M,$$

where s_i^2 is the realization of the random variable $S_i^2 = \sum_{s=1}^N U_{is}\sigma_s^2$. Also, due to independence of the permutations of units, $\text{cov}_{\xi_2|\xi_1}(\mathbf{Y}_i, \mathbf{Y}_{i^*}) = \mathbf{0}_{M\times M}$ for $i \neq i^*$. Combining these results for all $i = 1,\ldots,N$, $\text{var}_{\xi_2|\xi_1}(\mathbf{Y}_{NM\times 1}) = \mathbf{D}_\mathbf{s}\otimes\mathbf{P}_M$ where $\mathbf{D}_\mathbf{s}$ is a diagonal matrix with diagonal elements equal to the elements in the vector $\mathbf{s}_{N\times 1} = (s_i^2)$. Taking the expectation over random permutations of clusters, we have

$$E_{\xi_1}(S_i^2) = \sum_{s=1}^N \frac{\sigma_s^2}{N} = \sigma_e^2,$$

and, hence,

$$\text{var}_{\xi_1\xi_2}(\mathbf{Y}) = \sigma^2(\mathbf{P}_N\otimes\mathbf{J}_M) + \sigma_e^2(\mathbf{I}_N\otimes\mathbf{P}_M). \quad \text{(A.3)}$$

Adding (A.2) to (A.3) results in (24).

APPENDIX B: DEVELOPMENT OF THE EXPECTED MSE

We develop the expected MSE for the predictor given by (27) of the random variable

$$T = \begin{pmatrix}\mathbf{g}_I' & \mathbf{g}_{II}'\end{pmatrix}\begin{pmatrix}\mathbf{Y}_I\\ \mathbf{Y}_{II}\end{pmatrix}.$$

First, we express the predictor as

$$\hat{T} = (\mathbf{g}_I'\mathbf{A}' + \mathbf{g}_{II}'\mathbf{B}')\mathbf{Y}_I^*,$$

where

$$\mathbf{A}' = \left[\frac{\mathbf{J}_{nm}}{nm} + (1-\rho_t)k^*\left(\mathbf{P}_n\otimes\frac{\mathbf{J}_m}{m}\right) + \rho_t\mathbf{P}_{nm}\right]$$

and

$$\mathbf{B}' = \frac{\mathbf{J}_{(NM-nm)\times nm}}{nm} + k^*\begin{pmatrix}\left(\mathbf{P}_n\otimes\frac{\mathbf{J}_{(M-m)\times m}}{m}\right)\\ \hline \mathbf{0}_{(N-n)M\times nm}\end{pmatrix}.$$

As a result,

$$\hat{T} - T = \left((\mathbf{g}_I'\mathbf{A}' + \mathbf{g}_{II}'\mathbf{B}') \,\vdots\, -\mathbf{g}_I' \,\vdots\, -\mathbf{g}_{II}'\right)\begin{pmatrix}\mathbf{Y}_I^*\\ \hline \mathbf{Y}_I\\ \hline \mathbf{Y}_{II}\end{pmatrix}.$$

Using the expression for

$$\text{var}_{\xi_1\xi_2\xi_3}\begin{pmatrix}\mathbf{Y}_I^*\\ \hline \mathbf{Y}_I\\ \hline \mathbf{Y}_{II}\end{pmatrix} = \begin{pmatrix}\mathbf{V}_I^* & \mathbf{V}_I & \mathbf{V}_{I,II}\\ \mathbf{V}_I & \mathbf{V}_I & \mathbf{V}_{I,II}\\ \mathbf{V}_{II,I} & \mathbf{V}_{II,I} & \mathbf{V}_{II}\end{pmatrix},$$

the MSE simplifies to

$$MSE(\hat{T}) = (\mathbf{g}_I' \,\vdots\, \mathbf{g}_{II}')\begin{pmatrix}\mathbf{C}_{11} & \mathbf{C}_{21}'\\ \hline \mathbf{C}_{21} & \mathbf{C}_{22}\end{pmatrix}\begin{pmatrix}\mathbf{g}_I\\ \mathbf{g}_{II}\end{pmatrix}, \quad \text{(B.1)}$$

where $\mathbf{C}_{11} = (\mathbf{A}' - \mathbf{I}_{nm})\mathbf{V}_I(\mathbf{A} - \mathbf{I}_{nm}) + \sigma_r^2\mathbf{A}'\mathbf{A}$,

$$\mathbf{C}_{21} = (\mathbf{B}'\mathbf{V}_I - \mathbf{V}_{II,I})(\mathbf{A} - \mathbf{I}_{nm}) + \sigma_r^2\mathbf{B}'\mathbf{A},$$

and

$$\mathbf{C}_{22} = \mathbf{B}'\mathbf{V}_I\mathbf{B} - 2\mathbf{V}_{II,I}\mathbf{B} + \mathbf{V}_{II} + \sigma_r^2\mathbf{B}'\mathbf{B}.$$

Simplifying terms, we have

$$\mathbf{C}_{11} = \left[\rho_t\mathbf{I}_{nm} + (1-\rho_t)k^*\left(\mathbf{I}_n\otimes\frac{\mathbf{J}_m}{m}\right) - (1-\rho_t)(k^*-1)\frac{\mathbf{J}_{nm}}{nm}\right]\sigma_r^2,$$

$$\mathbf{C}_{21} = \begin{pmatrix}k^*\left(\mathbf{I}_n\otimes\frac{\mathbf{J}_{(M-m)\times m}}{m}\right) - (k^*-1)\frac{\mathbf{J}_{n(M-m)\times nm}}{nm}\\ \hline \frac{\mathbf{J}_{(N-n)M\times nm}}{nm}\end{pmatrix}\sigma_r^2,$$

and

$$\mathbf{C}_{22} = \begin{pmatrix}\sigma_e^2\mathbf{I}_{n(M-m)} + k^*(\sigma_e^2+\sigma_r^2)\left(\mathbf{P}_n\otimes\frac{\mathbf{J}_{(M-m)}}{m}\right)\\ + (\sigma_e^2+\sigma_r^2)\frac{\mathbf{J}_{n(M-m)}}{nm}\\ \hline (\sigma_e^2+\sigma_r^2)\frac{\mathbf{J}_{(N-n)M\times n(M-m)}}{nm}\end{pmatrix}$$

$$\vdots\begin{pmatrix}(\sigma_e^2+\sigma_r^2)\frac{\mathbf{J}_{nm\times(N-n)M}}{nm}\\ \hline \sigma_e^2\mathbf{I}_{(N-n)M} + m\sigma^{*2}\left(\mathbf{I}_{N-n}\otimes\frac{\mathbf{J}_M}{m}\right)\\ + (m\sigma^{*2}+\sigma_e^2+\sigma_r^2)\frac{\mathbf{J}_{(N-n)M}}{nm}\end{pmatrix}.$$

The expected MSE given by (32) is obtained by setting $\mathbf{g}'$ equal to (20) in (B.1) and simplifying. Expressions for the expected MSE of the predictor of T_A given by (29) can be developed in a similar manner.

[Received January 2003. Revised January 2004.]

REFERENCES

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

Cochran, W. (1977), *Survey Sampling*, New York: Wiley.

Godambe, V. P. (1955), "A Unified Theory of Sampling From Finite Populations," *Journal of the Royal Statistical Society*, Ser. B, 17, 269–278.

Goldberger, A. S. (1962), "Best Linear Unbiased Prediction in the Generalized Linear Regression Model," *Journal of the American Statistical Association*, 57, 369–375.

Graybill, F. A. (1983), *Matrices With Applications in Statistics*, Belmont, CA: Wadsworth International.

Harville, D. A. (1997), *Matrix Algebra From a Statistician's Perspective*, New York: Springer-Verlag.

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

Henderson, C. R., Kempthorne, O., Searle, S. R., and von Krosigk, C. M. (1959), "The Estimation of Environmental and Genetic Trends From Records Subject to Culling," *Biometrics*, 15, 192–218.

Hinkelmann, K., and Kempthorne, O. (1994), *Design and Analysis of Experiments*, *Vol. 1*, *Introduction to Experimental Design*, New York: Wiley.

Konijn, H. S. (1962), "Regression Analysis in Sample Surveys," *Journal of the American Statistical Association*, 57, 590–605.

Lencina, V. B. (2002), "Modelos de efeitos aleatorios e populacoes finites," unpublished doctoral thesis, University of Sao Paulo.

Li, W. (2003), "Use of Random Permutation Model in Rate Standardization and Calibration," unpublished doctoral thesis, University of Massachusetts, Massachusetts.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996), *SAS System for Mixed Models*, Cary, NC: SAS Institute.

McCulloch, C. E., and Searle, S. R. (2001), *Generalized, Linear, and Mixed Models*, New York: Wiley.

McLean, R. A., Sanders, W. L., and Stroup, W. W. (1991), "A Unified Approach to Mixed Linear Models," *The American Statistician*, 45, 54–64.

Merriam, P. A., Ockene, I. S., Hebert, J. R., Rosal, M. C., and Matthews, C. E. (1999), "Seasonal Variation of Blood Cholesterol Levels: Study Methodology," *Journal of Biologic Rhythms*, 14, 330–339.

Ockene, I. S., Chiriboga, D. E., Stanek, E. J., III, Harmatz, M. G., Nicolosi, R., Saperia, G., Well, A. D., Merriam, P. A., Reed, G., Ma, Y., Matthews, C. E., and Hebert, J. R. (2004), "Seasonal Variation in Serum Cholesterol: Treatment Implications and Possible Mechanisms," *Archives of Internal Medicine*, 164, 863–870.

Pfeffermann, D. (1984), "On Extensions of the Gauss–Markov Theorem to the Case of Stochastic Regression Coefficients," *Journal of the Royal Statistical Society*, Ser. B, 46, 139–148.

Pfeffermann, D., and Nathan, G. (1981), "Regression Analysis of Data From a Cluster Sample," *Journal of the American Statistical Association*, 76, 681–689.

Porter, R. M. (1973), "On the Use of Survey Sample Weights in the Linear Model," *Annals of Economic and Social Measurement*, 2, 141–158.

Rao, J. N. K., and Bellhouse, D. R. (1978), "Estimation of Finite Population Mean Under Generalized Random Permutation Model," *Journal of Statistical Planning and Inference*, 2, 125–141.

Robinson, G. K. (1991), "That BLUP Is a Good Thing: The Estimation of Random Effects," *Statistical Science*, 6, 15–51.

Royall, R. M. (1976), "The Linear Least-Squares Prediction Approach to Two-Stage Sampling," *Journal of the American Statistical Association*, 71, 657–664.

Scott, A., and Smith, T. M. F. (1969), "Estimation in Multi-Stage Surveys," *Journal of the American Statistical Association*, 64, 830–840.

Stanek, E. J., III, Singer, J. M., and Lencina, V. B. (2004), "A Unified Approach to Estimation and Prediction Under Simple Random Sampling," *Journal of Statistical Planning and Inference*, 121, 325–338.

Stanek, E. J., III, Well, A., and Ockene, I. (1999), "Why Not Routinely Use Best Linear Unbiased Predictors (BLUPs) as Estimates of Cholesterol, Percent Fat From Kcal and Physical Activity?" *Statistics in Medicine*, 18, 2943–2959.

Valliant, R., Dorfman, H. A., and Royall, R. M. (2000), *Finite Population Sampling and Inference*, New York: Wiley.

Willett, W. C. (1990), *Nutritional Epidemiology*, New York: Oxford University Press.