A unified approach to estimation and prediction under simple random sampling

Edward J. Stanek III[a],[*], Julio da Motta Singer[b],
Viviana Beatriz Lencina[c]

[a] *Department of Biostatistics and Epidemiology, SPH, University of Massachusetts at Amherst,*
404 Arnold House, 715 N. Pleasant Street, Amherst, MA 01003-9304, USA
[b] *Departamento de Estatística, IME, Universidade de São Paulo, Brazil*
[c] *Departamento de Investigación, FM, Universidad Nacional de Tucumán, Argentina*

Abstract

We consider a probability model where the design-based approach to inference under simple random sampling of a finite population encompasses a simple random permutation super-population model. The model consists of an expanded set of random variables following a random permutation probability distribution that keeps track of both the units' labels and positions in the permutation. In particular, since we keep track of the labels, the model allows us to attack the problem of estimation of a unit's parameter. While some linear combinations of the expanded set of random variables correspond to linear combinations of the unit parameters, other linear combinations correspond to random variables known as random effects. Using a prediction technique similar to that employed under the model-based approach, we develop optimum estimators of the linear combinations of the unit parameters and optimum predictors of the random effects.

The unbiased minimum variance estimator of the population mean is the sample mean and of a unit parameter is the Horvitz–Thompson estimator if the unit is included in the sample, and zero otherwise. The predictor of the random variable at a given position in the permutation is the realized unit's parameter for positions in the sample, and the sample mean for other positions. For other linear functions, unique minimum variance unbiased estimators may not exist.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Bias; Finite population; Optimal estimation; Prediction; Random effects; Mixed models;
Super-population; Design based; Model based; Inference

---

[*] Corresponding author. Fax: +1-413-5453812.
   E-mail address: stanek@schoolph.umass.edu (E.J. Stanek III).

1. Introduction

We propose a probability model induced by a simple random sample design for a finite population that encompasses a simple random permutation super-population model. Model-based prediction tools are used to optimally estimate linear combinations of random variables in the model. Appropriate linear combinations of the random variables may be constructed to represent finite population parameters, including the parameter for an individual unit. Other linear combinations correspond to random variables that are analogous to random effects. The model provides a common context for comparing results on prediction and estimation. Since the parameters can be estimated and the random variables can be predicted in a common manner, the results lead to interesting interpretations.

The probability model we propose was motivated by the desire to construct inference for a unit parameter in simple random sampling. While this problem is not of compelling interest, it is closely related to a similar common problem in two stage sampling where there is interest in predicting the parameter for a realized unit (or cluster). The investigation of that more complicated problem led us to focus on this simpler setting which still retains some essential aspects of the two stage problem. We explore the simpler setting here, deferring further comments on the two stage setting to the discussion.

Inference about a parameter for an individual labeled unit is not possible under the classical design-based approach since individual labeled units are not identifiable in the probability models generally used to link the sample to the population. In fact, the probability models used for such purposes are typically based on the distributions of exchangeable random variables which ignore labels. We overcome this problem by introducing a discrete probability model where parameters correspond to the values of the labeled units. The model is based on indicator random variables generated by a random permutation of units, as would occur in a simple random sampling design. These random variables keep track of both the unit's label and the unit's position in the permutation. Rather than characterizing a permuted finite population by N random variables, the expanded framework includes N^2 random variables.

The model we propose does not rely on the concept of a super-population considered under the model-based approach. However, estimators/predictors of linear combinations of random variables are constructed using the prediction approach common in model-based inference. Furthermore, linear combinations of the random variables reproduce the simple random permutation super-population model.

The problem we consider is particularly simple, and hence is related to a broad literature. The general modeling framework for survey sampling is given by Cassel et al. (1977), with design-based and model-based inference widely discussed (Bolfarine and Zacks, 1992; Hedayat and Sinha, 1991; Mukhopadhyay, 2001; Särndal et al., 1992; Thompson, 1997; Valliant et al., 2000). Recent reviews of inference in survey sampling are given by Rao (1997, 1999a). Brewer et al. (1988) and Brewer (1999, 2002) have discussed reconciling model-based and design-based inference. The random permutation super-population model has been discussed by Rao and Bellhouse (1978), Mukhopadhyay (1984) and Rao (1984), and in the context of two stage sampling, by

Padmawar and Mukhopadhyay (1985) and Bellhouse and Rao (1986). Model-based approaches to the two-stage problem have been studied by Scott and Smith (1969) and Fuller and Battese (1973), and recently reviewed in the context of small area estimation by Rao (1999b).

Of particular relevance are the fundamental results of Godambe (1955) and Godambe and Joshi (1965) that no uniform minimum variance unbiased linear estimator of the population total exists if coefficients are allowed to depend on the sequence of labels in the sample. Royall (1969) countered this result with the observation that if random variables representing the sampling were reduced to their usual representation, where one random variable is associated with each selected unit, optimal estimators could be obtained. Other approaches to overcome the non-existence result of Godambe have been suggested by Hartley and Rao (1968, 1969). Our approach is in the same spirit as that of Royall's 1969 result, where we reduce the most general set of random variables defined by Godambe to a set of N^2 random variables.

Definitions and notation are developed in Section 2 and the expanded model is fully defined in Section 3. Interest is focused on linear combinations of the random variables defined in the expanded model. Certain linear combinations simplify to non-stochastic finite population parameters; other linear combinations are random variables. Since both parameters and random variables can be defined by the linear combinations, the methods we develop in Section 4 are appropriate for both estimators (of parameters) and predictors (of random variables). For simplicity, we use the term 'estimator' in reference to general linear combinations of random variables.

The expanded model enables estimation of the population mean, as well as parameters for labeled units. The sample mean is the best linear unbiased estimate of the population mean. For a single unit, the best linear unbiased estimator is unique and of the Horvitz–Thompson (1952) type if the unit is included in the sample, and zero otherwise. Simultaneous estimation of all unit parameters in the population does not in general lead to unique estimators. However, with different additional restrictions, different unique estimators arise. The predictor of the random variable corresponding to the ith position in an ordered permutation, while not of any obvious interest, turns out to be analogous to the widely used predictor of a realized random effect in a mixed model. These results are discussed further in Section 5.

2. Definitions and notation

We consider the problem of estimating certain characteristics of a finite population of units under simple random without replacement sampling. We define a finite population as a collection of a known number, N, of identifiable units labeled $j = 1, \ldots, N$. Associated with unit j is a parameter y_j. We summarize the set of parameters in the vector $\mathbf{y} = (y_1, \ldots, y_N)'$ and assume that when unit j is observed, the parameter y_j is known without error. Typically, there is interest in a $p \times 1$ vector of parameters of the form $\beta = \mathbf{G}\mathbf{y}$ where $\mathbf{G}$ is a matrix of known constants. For example, if $\mathbf{G} = \mathbf{I}_N$, with $\mathbf{I}_N$ denoting the N-dimensional identity matrix, then β is the set of individual parameters. If $\mathbf{G} = \mathbf{e}_j'$, where $\mathbf{e}_j$ denotes an N-dimensional column vector with null elements in all

positions except for the jth position for which the value 1 is assigned, the parameter β corresponds to the value y_j associated with the unit labeled j in the population. When $\mathbf{G} = N^{-1}\mathbf{1}'_N$, where $\mathbf{1}_N$ denotes an N-vector with all elements equal to 1, β corresponds to the population mean, μ.

We define a probability model that links the population parameters to an expanded vector of random variables which is essentially induced by a simple random sampling design, and develop estimators of linear functions of these random variables. The proposed estimators are linear functions of the random variables that define a sample. We use the prediction approach that is common in model-based inference to develop the estimators. Before introducing the expanded model, we first review the prediction approach used in the context of super-population models.

The prediction approach is based on an underlying probability model for a vector of random variables $\mathbf{Y}^* = (Y_1, \ldots, Y_N)'$ that characterizes a super-population. The population under study, $\mathbf{y} = (y_1, \ldots, y_N)'$, is considered to be a realization of these super-population random variables. The vector of random variables is partitioned into a subset which we call the sample, $\mathbf{Y}^*_S = (Y_1, \ldots, Y_n)'$ and the remainder, $\mathbf{Y}^*_R = (Y_{n+1}, \ldots, Y_N)'$, such that $\mathbf{Y}^* = (\mathbf{Y}^{*\prime}_S, \mathbf{Y}^{*\prime}_R)'$. Inference is solely based on linear models of the form

$$\mathbf{Y}^* = \mathbf{X}^*\beta^* + \mathbf{E}^*, \tag{2.1}$$

where $\mathbf{X}^*$ is a known non-stochastic matrix, β^* is a p-dimensional vector of super-population parameters and $\mathbf{E}^*$ is an N-dimensional vector of random errors governed by the probability model under which $E_\xi(\mathbf{E}^*) = \mathbf{0}$, where ξ denotes expectation with respect to the super-population. Although the super-population parameters appear in the model, they are not of primary interest. Instead, the parameters of interest are linear combinations $\beta = \mathbf{Gy}$ of a realization of $\mathbf{Y}^*$. The population mean and the population total are typical examples of β.

Assuming that $\mathbf{Y}^*_S$ is realized, the estimator of β is based on the predictor of $\mathbf{Y}^*_R$ or some functions of it, in such a way that it satisfies some optimality criteria (see Royall (1976) or Bolfarine and Zacks (1992), for example). More specifically, Valliant et al. (2000, pp. 29–30) point out that the target parameters may be written as $\beta = \beta_S + \beta_R$, where β_S denotes the part of the linear combination observed in the sample and β_R denotes the part associated with the non-sampled units. After selecting the sample, the problem of estimating β is equivalent to predicting β_R and the best linear unbiased estimate (BLUE) of β is obtained by adding the optimal predictor of β_R to β_S. The prediction process relies on the probability model for the super-population and does not necessarily depend on the physical process used to select the sample.

3. The expanded model

Our main objective is to express an expanded set of random variables induced by the design-based approach in the form of model (2.1). We show that this model allows the construction of estimators of linear combinations of the corresponding random variables based on the same optimality criteria considered under the prediction approach. Some of these linear combinations correspond to population parameters, while others are

random variables. We restrict ourselves to the case where the sample is selected by simple random sampling without replacement. We first describe the typical design-based random permutation model, and then introduce the expanded model. An advantage of the expanded model is the ability to identify a parameter associated with a labeled unit.

Assuming simple random without replacement sampling, the typical random permutation probability model assigns equal probability to all permutations of the finite population units. We index each unit's position in the permutation by $i = 1, \ldots, N$. The value in position i for a randomly selected permutation is defined by the realization of the random variable $\tilde{Y}_i = \sum_{j=1}^{N} U_{ij} y_j$ where $U_{ij} = 1$ if unit j is in position i and $U_{ij} = 0$ otherwise. The random vector $\tilde{\mathbf{Y}} = (\tilde{Y}_1 \ \tilde{Y}_2 \ \cdots \ \tilde{Y}_N)'$ is the random permutation super-population (Cassel et al., 1977), and the random variables $\tilde{Y}_i$, $i = 1, \ldots, n$, correspond to a sample. This representation of random variables does not allow units to be identified and hence does not permit inference about unit parameters.

The expanded model is based on representing the random variables in the sum $\sum_{j=1}^{N} U_{ij} y_j$ as individual random variables of the form $Y_{ij} = U_{ij} y_j$, which we summarize in an $N^2 \times 1$ vector $\mathbf{Y} = (\mathbf{Y}_1' \ \mathbf{Y}_2' \ \cdots \ \mathbf{Y}_N')'$ where $\mathbf{Y}_j = (Y_{1j} \ Y_{2j} \ \cdots \ Y_{Nj})'$. The vector of random variables can be defined compactly as $\mathbf{Y} = (\mathbf{D_y} \otimes \mathbf{I}_N) \operatorname{vec}(\mathbf{U})$, where $\otimes$ denotes the Kronecker product (Searle, 1982), $\mathbf{D_y}$ is a diagonal matrix with the elements of $\mathbf{y}$ along the main diagonal, $\operatorname{vec}(\mathbf{U})$ is a vector representing the column expansion of $\mathbf{U}$, and

$$\mathbf{U} = \begin{pmatrix} U_{11} & U_{12} & \ldots & U_{1N} \\ U_{21} & U_{22} & \ldots & U_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ U_{N1} & U_{N2} & \cdots & U_{NN} \end{pmatrix}.$$

Given the random structure of $\mathbf{U}$, the expected value and the variance of the expanded random vector are respectively given by

$$E(\mathbf{Y}) = \mathbf{Xy} \tag{3.1}$$

and

$$\operatorname{var}(\mathbf{Y}) = \mathbf{\Delta} \otimes \mathbf{P}_N, \tag{3.2}$$

where $\mathbf{X} = \mathbf{I}_N \otimes \mathbf{1}_N / N$, $\mathbf{P}_a = \mathbf{I}_a - a^{-1} \mathbf{J}_a$ with $\mathbf{J}_a = \mathbf{1}_a \mathbf{1}_a'$, and

$$\mathbf{\Delta} = \frac{1}{N-1} \mathbf{D_y} \mathbf{P}_N \mathbf{D_y}. \tag{3.3}$$

The selection of a simple random sample of size n from the population will result in the realization of nN of the expanded random variables in the vector $\mathbf{Y}$. We gather these random variables for the sample in the vector $\mathbf{Y}_S = (\oplus_{j=1}^{N} (\mathbf{I}_n \mid \mathbf{0}_{n \times (N-n)})) \mathbf{Y}$ by rearranging the elements in the vector $\mathbf{Y}$; similarly, the remaining $(N-n)N$ random variables are defined by the vector $\mathbf{Y}_R = (\oplus_{j=1}^{N} (\mathbf{0}_{(N-n) \times n} \mid \mathbf{I}_{(N-n)})) \mathbf{Y}$, where $\oplus_{j=1}^{N} \mathbf{A}_j$ denotes a block diagonal matrix, with blocks given by $\mathbf{A}_j$ (Searle, 1982). The variance

of the rearranged expanded random vector is partitioned as

$$\mathrm{Var}\begin{pmatrix} \mathbf{Y}_S \\ \mathbf{Y}_R \end{pmatrix} = \begin{pmatrix} \mathbf{V}_S & \mathbf{V}_{SR} \\ \mathbf{V}_{RS} & \mathbf{V}_R \end{pmatrix},$$

where $\mathbf{V}_S = \boldsymbol{\Delta} \otimes (\mathbf{I}_n - N^{-1}\mathbf{J}_n)$ and $\mathbf{V}_{SR} = \boldsymbol{\Delta} \otimes (-N^{-1}\mathbf{J}_{n \times (N-n)})$, with $\mathbf{J}_{n \times (N-n)} = \mathbf{1}_n \mathbf{1}'_{N-n}$.

As an illustration, consider a finite population with $N=4$ units from which we select a simple random sample without replacement of size $n=2$. Letting $\mathbf{y} = (y_1 \ y_2 \ y_3 \ y_4)'$ it follows that $\mathbf{Y} = (y_1(U_{11} \ U_{21} \ U_{31} \ U_{41}) \mid y_2 \ (U_{12} \ U_{22} \ U_{32} \ U_{42}) \mid y_3 \ (U_{13} \ U_{23} \ U_{33} \ U_{43}) \mid y_4 \ (U_{14} \ U_{24} \ U_{34} \ U_{44}))'$, $\mathbf{Y}_S = (y_1(U_{11} \ U_{21}) \mid y_2(U_{12} \ U_{22}) \mid y_3(U_{13} \ U_{23}) \mid y_4(U_{14} \ U_{24}))'$ and $\mathbf{Y}_R = (y_1(U_{31} \ U_{41}) \mid y_2(U_{32} \ U_{42}) \mid y_3(U_{33} \ U_{43}) \mid y_4(U_{34} \ U_{44}))'$. Supposing that the first and second selected units in a sample are units 3 and 1, respectively, the realized value of $\mathbf{Y}_S$ is $(0 \ y_1 \mid 0 \ 0 \mid y_3 \ 0 \mid 0 \ 0)'$.

4. Estimation

A characteristic of the proposed model is that the vector of parameters $\mathbf{y}$ may be defined as linear combinations $\mathbf{LY}$ of the expanded random variables. For example, setting

$$\mathbf{L} = \mathbf{I}_N \otimes \mathbf{1}'_N, \tag{4.1}$$

$\mathbf{LY} = \mathbf{y}$, while the value for unit j in the population, y_j, is defined by setting

$$\mathbf{L} = \mathbf{e}'_j \otimes \mathbf{1}'_N. \tag{4.2}$$

The population mean μ is defined by setting

$$\mathbf{L} = N^{-1}\mathbf{1}'_{N^2}. \tag{4.3}$$

More generally, we can define other linear combinations of $\mathbf{Y}$ which are stochastic. For example, a random variable corresponding to the value that will appear in the ith position in a permutation is defined by setting

$$\mathbf{L} = \mathbf{1}'_N \otimes \mathbf{e}'_i. \tag{4.4}$$

In general, for linear combinations defined in terms of the expanded random variables, we can discuss estimating a parameter or predicting a random variable. The specification of $\mathbf{L}$ is necessary to determine whether $\mathbf{LY}$ is fixed or random. We can encompass both the estimation and the prediction problems in the same framework. As previously noted for simplicity, we use the term 'estimation' in reference to a general linear combination of random variables.

It is not necessary to use the expanded random variables to develop estimators for all linear combinations of $\mathbf{Y}$. To see this, we evaluate the linear combination using the expansion given by

$$\mathbf{Y} = \left(\frac{\mathbf{1}_N}{N} \otimes \mathbf{I}_N\right)\tilde{\mathbf{Y}} + (\mathbf{P}_N \otimes \mathbf{I}_N)\mathbf{Y}. \tag{4.5}$$

For example, using (4.3), the linear combination defining the population mean simplifies to $\mathbf{LY} = (1/N)\mathbf{1}'_N\tilde{\mathbf{Y}}$. Similarly, using (4.4), the linear combination defining the random

variable corresponding to the value that will appear in the ith position in a permutation simplifies to $\mathbf{LY} = \mathbf{e}_i' \tilde{\mathbf{Y}}$. For the linear combinations defined by (4.3) and (4.4), the optimal estimator can be developed by solely considering the random variables $\tilde{\mathbf{Y}}$ since the first and second terms in (4.5) are orthogonal, and the second term has expected value equal to zero (Rao and Bellhouse, 1978, Theorem 1.1).

Using the prediction approach, we develop the solution to the problem of estimating $\mathbf{LY}$ based on a sample. First, we partition $\mathbf{LY}$ into a sample component, $\mathbf{L}_S \mathbf{Y}_S$, and a remaining component, $\mathbf{L}_R \mathbf{Y}_R$. We require the predictors of $\mathbf{L}_R \mathbf{Y}_R$ to be linear in the sample, and represent them by $\mathbf{L}_{RS} \mathbf{Y}_S$ Defining $\mathbf{C} = \mathbf{L}_S + \mathbf{L}_{RS}$, the class of estimators of $\mathbf{LY}$ is given by

$$C_E = \{\mathbf{CY}_S: \mathbf{C} \text{ is a } p \times Nn \text{ matrix of constants}\}.$$

We require the estimators to be unbiased (such that $E(\mathbf{CY}_S - \mathbf{LY}) = \mathbf{0}$), and have minimum generalized mean squared error given by

$$\text{GMSE} = \text{Var}[\mathbf{1}_p'(\mathbf{CY}_S - \mathbf{LY})] \tag{4.6}$$

(Bolfarine and Zacks, 1992). Using (3.1), we may write $E(\mathbf{CY}_S) = \mathbf{CX}_S \mathbf{y}$, where

$$\mathbf{X}_S = N^{-1} \mathbf{I}_N \otimes \mathbf{1}_n \tag{4.7}$$

so that the unbiased condition reduces to $\mathbf{CX}_S \mathbf{y} = \mathbf{LX} \mathbf{y}$ for all $\mathbf{y}$, or equivalently

$$\mathbf{CX}_S = \mathbf{LX}. \tag{4.8}$$

We solve (4.8) for $\mathbf{C}$ in terms of an arbitrary matrix, and then minimize the GMSE with respect to that matrix. When $\mathbf{LY}$ is non-stochastic, the result is given by

$$\hat{\mathbf{C}} = \mathbf{L}\left(\mathbf{I}_N \otimes \frac{\mathbf{J}_{N \times n}}{n}\right) + \mathbf{P}_p \mathbf{T}_{p \times nN}'(\mathbf{I}_N \otimes \mathbf{P}_n), \tag{4.9}$$

where $\mathbf{T}'$ is an arbitrary matrix resulting from use of generalized inverses to obtain the solution (see Appendix A).

Solutions to the problem of estimating a linear function of $\tilde{\mathbf{L}}\tilde{\mathbf{Y}}$ are developed in a similar manner. We briefly outline the solution that was first given by Royall (1976). First, note that $E(\tilde{\mathbf{Y}}) = \tilde{\mathbf{X}}\mu$, where $\tilde{\mathbf{X}} = \mathbf{1}_N$, and $\text{Var}(\tilde{\mathbf{Y}}) = \sigma^2 \mathbf{P}_N$, where $\sigma^2 = (1/(N-1)) \sum_{j=1}^N (y_j - \mu)^2$. Partitioning $\tilde{\mathbf{Y}}$ into the sample, $\tilde{\mathbf{Y}}_S = (\tilde{Y}_1, \ldots, \tilde{Y}_n)'$, and the remainder, $\tilde{\mathbf{Y}}_R = (\tilde{Y}_{n+1}, \ldots, \tilde{Y}_N)'$, results in

$$\text{Var}\begin{pmatrix} \tilde{\mathbf{Y}}_S \\ \tilde{\mathbf{Y}}_R \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{V}}_S & \tilde{\mathbf{V}}_{SR} \\ \tilde{\mathbf{V}}_{RS} & \tilde{\mathbf{V}}_R \end{pmatrix},$$

where $\tilde{\mathbf{V}}_S = \sigma^2(\mathbf{I}_n - N^{-1}\mathbf{J}_n)$ and $\tilde{\mathbf{V}}_{SR} = -\sigma^2(N^{-1}\mathbf{J}_{n \times (N-n)})$. We partition $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{L}}$ in a similar manner resulting in $\tilde{\mathbf{X}}_S = \mathbf{1}_n$, $\tilde{\mathbf{X}}_R = \mathbf{1}_{N-n}$ and $\tilde{\mathbf{L}}\tilde{\mathbf{Y}} = \tilde{\mathbf{L}}_S \tilde{\mathbf{Y}}_S + \tilde{\mathbf{L}}_R \tilde{\mathbf{Y}}_R$. We require the predictor of $\tilde{\mathbf{L}}_R \tilde{\mathbf{Y}}_R$ to be a linear function of the sample, $\tilde{\mathbf{L}}_{RS} \tilde{\mathbf{Y}}_S$, to be unbiased, i.e. to satisfy $E(\tilde{\mathbf{L}}_{RS} \tilde{\mathbf{Y}}_S) = E(\tilde{\mathbf{L}}_R \tilde{\mathbf{Y}}_R)$, and to have minimum GMSE (given by

$\mathrm{var}[\mathbf{1}_p'(\tilde{\mathbf{C}}\tilde{\mathbf{Y}}_S - \tilde{\mathbf{L}}\tilde{\mathbf{Y}})]$, where $\tilde{\mathbf{C}} = \tilde{\mathbf{L}}_S + \tilde{\mathbf{L}}_{RS}$). The resulting estimator is

$$\hat{\tilde{\mathbf{C}}}\tilde{\mathbf{Y}}_S = \tilde{\mathbf{L}}_S\tilde{\mathbf{Y}}_S + \tilde{\mathbf{L}}_R[\tilde{\mathbf{X}}_R\hat{\alpha} + \tilde{\mathbf{V}}_{SR}'\tilde{\mathbf{V}}_S^{-1}(\tilde{\mathbf{Y}}_S - \tilde{\mathbf{X}}_S\hat{\alpha})], \tag{4.10}$$

where $\hat{\alpha} = (\tilde{\mathbf{X}}_S'\tilde{\mathbf{V}}_S^{-1}\tilde{\mathbf{X}}_S)^{-1}\tilde{\mathbf{X}}_S'\tilde{\mathbf{V}}_S^{-1}\tilde{\mathbf{Y}}_S$.

4.1. Estimating y_j

We obtain the estimator of $\mathbf{LY}$ with $\mathbf{L}$ defined by (4.2) corresponding to a particular value y_j associated with the unit labeled j. Since $p = 1$, $\mathbf{P}_p = 0$ and (4.9) simplifies to

$$\hat{\mathbf{C}} = \frac{N}{n}(\mathbf{e}_j' \otimes \mathbf{1}_n'). \tag{4.11}$$

This corresponds to $(N/n)y_j$ when unit j is included in the sample, and zero otherwise, a Horvitz–Thompson type estimator of the unit's value. For such an estimator, GMSE $= ((N-n)/n)y_j^2$.

4.2. Estimating y

We develop simultaneous unbiased estimators of all the individual parameters, $\mathbf{y}$, in a finite population next. These parameters are defined by setting $\mathbf{L}$ equal to (4.1). Since $p = N$, the solution given by (4.9) simplifies to

$$\hat{\mathbf{C}} = \frac{N}{n}(\mathbf{I}_N \otimes \mathbf{1}_n') + \mathbf{P}_N\mathbf{T}_{N \times nN}'(\mathbf{I}_N \otimes \mathbf{P}_n), \tag{4.12}$$

where $\mathbf{T}$ is an arbitrary matrix. In general, the second term in (4.12) is not zero, and hence there are multiple solutions, each of which has GMSE $= (N(N-n)/n)\sigma^2$.

Unique estimators can be obtained by imposing restrictions on the structure of the coefficients, $\mathbf{C}$. For example, if we assume that $\mathbf{C} = \mathbf{I}_N \otimes \mathbf{v}_{1 \times n}'$, where $\mathbf{v}$ is a vector of unknown constants, following the same strategy, we may show that the unique estimator of $\mathbf{y}$ is

$$\hat{\mathbf{C}}\mathbf{Y}_S = \frac{N}{n}(\mathbf{I}_N \otimes \mathbf{1}_n)\mathbf{Y}_S. \tag{4.13}$$

This restriction forces the coefficients to be the same for different parameters, but allows the coefficients to differ with position. However, not all structures for $\mathbf{C}$ lead to unbiased estimators. For example, there are no solutions for $\mathbf{C} = \mathbf{J}_N \otimes \mathbf{v}_{1 \times n}'$.

A more general class of estimators can be considered if we replace the requirement of unit unbiasedness by average unbiasedness, $\mathbf{1}_N'E(\mathbf{CY}_S - \mathbf{LY}) = 0$. With this requirement, and proceeding in a manner similar to that used to obtain (4.12), the estimator of $\mathbf{y}$ simplifies to

$$\hat{\mathbf{C}}\mathbf{Y}_S = \frac{1}{n}\mathbf{J}_{N \times nN}\mathbf{Y}_S + \mathbf{P}_N\mathbf{T}_{N \times nN}'\mathbf{Y}_S. \tag{4.14}$$

This estimator is not unique since $\mathbf{T}$ is arbitrary. If $\mathbf{C} = \mathbf{I}_N \otimes \mathbf{v}_{1 \times n}'$, a unique solution results and is given by (4.13). If $\mathbf{C}$ is restricted to be of the form $\mathbf{C} = \mathbf{J}_N \otimes \mathbf{v}_{1 \times n}'$, it follows that the unique solution is $\hat{\mathbf{C}}\mathbf{Y}_S = \bar{y}\mathbf{1}_N$, the sample mean for each element.

We illustrate these results via a simple example. Let us assume that $N = 4$, $n = 2$, $\mathbf{y} = (y_1 \ y_2 \ y_3 \ y_4)'$ and that the realized value of $\mathbf{Y}_S$ is $(0 \ y_1 \ | \ 0 \ 0 \ | \ y_3 \ 0 \ | \ 0 \ 0)'$, i.e., the third unit was selected in the first position and the first unit was selected in the second position in the sample. If we require the estimator (4.12) to be linear and unbiased, only the estimator for the unselected unit (and also for units j for which $y_j = 0$) is unique, and equal to zero. The estimate for unit $j = 1$ is $a y_1$, while the estimate for unit $j = 3$ is $c y_3$ with a and c denoting functions of elements in the arbitrary matrix, $\mathbf{T}$. If we require the estimator to be linear and unbiased, and restrict the coefficients to be of the form $\mathbf{C} = \mathbf{I}_N \otimes \mathbf{v}'_{1 \times n}$, then the unique estimates for units $j = 1$ and $j = 3$ are given by the Horvitz–Thompson type estimate, $\frac{4}{2} y_j$. The estimates for unit $j = 2$ and $j = 4$ are zero. Using the average unbiased constraint, and requiring estimators to be linear in the sample with coefficients of the form $\mathbf{C} = \mathbf{J}_N \otimes \mathbf{v}'_{1 \times n}$, the unique estimate for all units is given by the sample mean, $\bar{y}$.

4.3. Estimating μ and predicting random variable(s) in the ith position in a permutation based on $\tilde{\mathbf{Y}}$

The linear combination $\mathbf{LY}$ with $\mathbf{L}$ given by (4.3) defines the population mean; setting $\mathbf{L}$ equal to (4.4) defines the random variable that will appear in the ith position in a permutation. Using (4.5), both linear combinations are equal to linear functions of $\tilde{\mathbf{L}}\tilde{\mathbf{Y}}$ with $\tilde{\mathbf{L}} = \mathbf{1}'_N/N$, and $\tilde{\mathbf{L}} = \mathbf{e}'_i$, respectively. Using the coefficients that define the population mean, and noting that $\tilde{\mathbf{V}}_S^{-1} = (1/\sigma^2)(\mathbf{I}_n + \mathbf{J}_n/(N-n))$ and $\hat{\alpha} = N\bar{y}$, estimator (4.10) simplifies to $\bar{y}$, the sample mean.

We partition $\mathbf{e}_i = (\mathbf{e}'_{iS} \ \mathbf{e}'_{iR})'$ where $\mathbf{e}_{iS}$ is a vector of dimension $n \times 1$ to predict the random variable in the ith position in a permutation. When $i \leqslant n$, $\hat{\tilde{\mathbf{C}}}\tilde{\mathbf{Y}}_S = \mathbf{e}'_{iS}\tilde{\mathbf{Y}}_S$ which will correspond to the value of the unit that is in the ith position in a realized permutation, i.e. $\sum_{j=1}^{N} u_{ij} y_j$ (where u_{ij} represents the realized value of U_{ij}). When $i > n$, $\tilde{\mathbf{L}}\tilde{\mathbf{Y}} = \tilde{\mathbf{L}}_R\tilde{\mathbf{Y}}_R$, and $\hat{\tilde{\mathbf{C}}}\tilde{\mathbf{Y}}_S = \mathbf{e}'_{iR}[\tilde{\mathbf{X}}_R\hat{\alpha} + \tilde{\mathbf{V}}'_{SR}\tilde{\mathbf{V}}_S^{-1}(\tilde{\mathbf{Y}}_S - \tilde{\mathbf{X}}_S\hat{\alpha})]$ which simplifies to $\bar{y}$. The GMSE of the predictor is zero when $i \leqslant n$, and equal to $\sigma^2((n+1)/n)$ when $i > n$.

Simultaneous predictors of the units realized in all N positions are defined by setting $\tilde{\mathbf{L}} = \mathbf{I}_N$ and result in the same predictors as those obtained for the individual positions. The predictors correspond to the realized unit's values when $i \leqslant n$, and to the sample mean when $i > n$. For the vector of predictors, $\mathrm{GMSE} = \sigma^2 N(N-n)/n$, which is equal to the GMSE of the estimator of $\mathbf{y}$ in Section 4.2.

5. Discussion

Design-based and model-based methods are usually discussed as separate approaches for estimation and inference in finite population sampling. We have presented an expanded probability model induced by the possible physical process of simple random sampling. Since no super-population model is required and the probability model arises solely from sampling, we consider the resulting estimators to be design-based. No

additional assumptions or concepts are required for estimation, which is accomplished by developing predictors of linear functions of the unobserved random variables. Linear functions of the expanded probability model lead to a set of random variables referred to by others as a simple random permutation super-population model (Cassel et al., 1977). Thus, the expanded model encompasses both design- and model-based frameworks. Although we feel that the expanded model unifies aspects of survey sampling methodology for simple random sampling, it has not yet been extended to the broad class of super-population models, including the more general random permutation super-population models.

Others have investigated a random permutation model in the context of a super-population framework, and concluded that the sample mean is the uniform minimum variance unbiased estimator of the population mean (Rao and Bellhouse, 1978). In such a framework, the likelihood is uninformative for unit parameters, and estimation has focused on the mean. Inclusion probabilities for labeled units, as opposed to the basic indicator random variables underlying unit selection are used. Although the indicator random variables used to define the expanded probability model are not new (see, for example, (Neyman, 1934; Neyman et al., 1935; Kempthorne, 1952)), their use in developing estimators of unit parameters appears to be novel.

The expanded model extends the typical permutation model to a broader set of random variables, but falls short of the very general set of random variables envisioned by Godambe (1955) which spans an $(N-1)^n$ dimensional space. The random variables in a typical permutation model span an $N-1$ dimensional space. The random variables in the expanded model span an $(N-1)^2$ dimensional space. Higher dimensional random variables may be postulated intermediate to Godambe's general model that may lead to new insights.

Our motivation in developing the expanded permutation model was to improve our understanding of realized random effects in the context of a mixed model. In a mixed model, a realized random effect is commonly defined as the difference between the parameter for a realized unit, and the mean of a population. With this definition, the expected value of a random effect is zero. To simplify the discussion, we define a realized random effect as the parameter for a labeled unit that is realized at a particular position in a permutation. Our definition is a re-parameterization of the definition commonly used for mixed models.

If a unit is included in a simple random sample, the realized random effect is simply the parameter for that unit. The value of the parameter (which is observed) is the best linear unbiased predictor. Since the predictor is the parameter for the realized unit, may we interpret the predictor as a predictor of the parameter for a specified unit? The expanded model provides the answer to this question since we can predict a random effect and a specified unit as separate linear combinations of random variables in the same model. The linear combinations that define these two quantities differ, as do their estimators. A clearer statement of the interpretation for what is commonly referred to as "the predictor of a realized random effect" is the predictor of a position in a permutation. In fact, since the expected value of the random variable at a position is the population mean, the predictor of this parameter will almost never equal the parameter being predicted.

In an analogous manner, the predictor of a 'realized random effect' in a simple mixed model will carry the interpretation as the predictor of the expected value of units that can occur at a position in a permutation. Similar results based on an expanded model for cluster sampling, while outside the scope of this paper, have been developed for equal size clusters both with and without response error (Stanek and Singer, 2002a) and in an unbalanced setting (Stanek and Singer, 2002b). The expanded framework is particularly important to retain the nesting of secondary sampling units in primary sampling units in an unbalanced two stage sampling context. Such results share the basic awkward interpretation as predictors of positions, not identified units, as illustrated in the expanded simple random sampling model.

While the results presented here are for simple random sampling, extensions to many other sample settings appear to be feasible. Such extensions include adding measurement error to simple random sampling, stratified sampling, and unbalanced cluster sampling settings. Strategies that account for covariates have been initially addressed in dissertations by Lencina (2002) and Li (2003). Extensions also appear feasible for experimental studies. There are also limitations. The two stage sample results are limited by the current lack of an optimal strategy for variance component estimation. Strategies for handling a continuous covariate are not yet developed and may not be feasible. Extensions to unequal probability sampling, may be possible but have not yet been developed.

From a different perspective, in the expanded model, linear combinations of random variables that correspond to unit parameters can be defined, and have a clear interpretation. The unbiased estimator of a unit's parameter (which corresponds to the Horvitz–Thompson estimator when the unit is included in the sample, or zero otherwise) suffers from the criticism of Basu's (1971) elephant example. The estimator is not intuitive, although it clearly satisfies the constraint for unbiasedness.

Many practitioners have used the predictor of a position in a permutation as an estimate of the parameter for a unit in the population. Such an estimator corresponds to the value for the unit if it is included in the sample or to the sample mean if it is not in the sample and may be written as

$$\hat{y}_s = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n} [I_{\{j=s\}} Y_{ij} + (1 - I_{\{j=s\}}) Y_{ij}/n] \right\},$$

where $I_{\{j=s\}}$ denotes an indicator function. This ad hoc estimator may be expressed in terms of the elements of the expanded random vector $\mathbf{Y}$ but not in terms of the collapsed random variables $\tilde{Y}_i$. However, it is a non-linear function of $\mathbf{Y}$, suggesting that beyond the need of keeping track of both labels and values attached to the units in the population for which we want to draw inference, a broader class of estimators is needed to obtain such a result. One way the non-linearity can be avoided is to define an extended set of random variables, beyond those proposed in this paper. Current research is underway to investigate such expanded sets, and use them to develop linear predictors of specific units.

Appendix A. Optimal estimators

We solve (4.8) for $\mathbf{C}'$ and then minimize the GMSE with respect to that matrix. First, note that (4.8) can be re-expressed as $\mathbf{X}'_S\mathbf{C}' = \mathbf{X}'\mathbf{L}'$. In general for fixed matrices $\mathbf{A}$ and $\mathbf{B}$ the set of solutions to $\mathbf{AW} = \mathbf{B}$ is given by $\mathbf{W} = \mathbf{A}^-\mathbf{B} + (\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{Z}$ where $\mathbf{A}^-$ is a specific g-inverse of $\mathbf{A}$ and $\mathbf{Z}$ is an arbitrary matrix (as defined by Graybill, 1983). We make repeated use of this result in obtaining the solution. Setting $\mathbf{X}'^-_S = \mathbf{I}_N \otimes N\mathbf{1}_n/n$, all solutions that satisfy the constraint for unbiasedness are given by

$$\mathbf{C}' = \left(\mathbf{I}_N \otimes \frac{\mathbf{J}_{n \times N}}{n}\right)\mathbf{L}' + (\mathbf{I}_N \otimes \mathbf{P}_n)\mathbf{Z}, \tag{A.1}$$

where $\mathbf{Z}_{nN \times p}$ is an arbitrary matrix. When $\mathbf{LY}$ is non-stochastic, (as in (4.1), (4.2) and (4.3)), the GMSE in (4.6) simplifies to

$$\text{GMSE} = \mathbf{1}'_p\text{Var}(\mathbf{CY}_S)\mathbf{1}_p = \mathbf{1}'_p\mathbf{CV}_S\mathbf{C}'\mathbf{1}_p,$$

which is a function of $\mathbf{Z}$. Defining

$$\mathbf{c} = \mathbf{Z}\mathbf{1}_p \tag{A.2}$$

and $\mathbf{a} = (\mathbf{I}_N \otimes \mathbf{J}_{n \times N}/n)\mathbf{L}'\mathbf{1}_p$, the GMSE simplifies to

$$\text{GMSE} = \mathbf{a}'\mathbf{V}_S\mathbf{a} + \mathbf{c}'(\boldsymbol{\Delta} \otimes \mathbf{P}_n)\mathbf{c} + 2\mathbf{c}'(\boldsymbol{\Delta} \otimes \mathbf{P}_n)\mathbf{a}. \tag{A.3}$$

Differentiating (A.3) with respect to $\mathbf{c}$ and setting the resulting derivatives equal to zero yields $(\boldsymbol{\Delta} \otimes \mathbf{P}_n)\hat{\mathbf{c}} = -(\boldsymbol{\Delta} \otimes \mathbf{P}_n)\mathbf{a}$. Since $\mathbf{P}_n$ is orthogonal to $\mathbf{J}_{n \times N}$, $(\boldsymbol{\Delta} \otimes \mathbf{P}_n)\mathbf{a} = \mathbf{0}$ and the solutions are given as

$$\hat{\mathbf{c}} = [\mathbf{I}_{nN} - (\boldsymbol{\Delta}^-\boldsymbol{\Delta} \otimes \mathbf{P}_n^-\mathbf{P}_n)]\mathbf{r}_{nN \times 1}, \tag{A.4}$$

where $\mathbf{r}$ is an arbitrary vector.

We replace $\mathbf{c}$ by (A.4) in Eq. (A.2), and solving for $\hat{\mathbf{Z}}'$, results in

$$\hat{\mathbf{Z}}'_{p \times nN} = \frac{1}{p}\mathbf{1}_p\mathbf{r}'[\mathbf{I}_{nN} - (\boldsymbol{\Delta}\boldsymbol{\Delta}^- \otimes \mathbf{P}_n\mathbf{P}_n^-)] + \mathbf{P}_p\mathbf{T}'_{p \times nN}, \tag{A.5}$$

where $\mathbf{1}'_p{}^- = \mathbf{1}_p/p$, and $\mathbf{T}'$ is an arbitrary matrix. Substituting (A.5) into (A.1) and simplifying,

$$\hat{\mathbf{C}} = \mathbf{L}\left(\mathbf{I}_N \otimes \frac{\mathbf{J}_{N\times n}}{n}\right) + \frac{1}{p}\mathbf{1}_p\mathbf{r}'_{1\times nN}[(\mathbf{I}_N - \boldsymbol{\Delta}\boldsymbol{\Delta}^-) \otimes \mathbf{P}_n]$$
$$+ \mathbf{P}_p\mathbf{T}'_{p\times nN}(\mathbf{I}_N \otimes \mathbf{P}_n), \qquad (A.6)$$

where both $\mathbf{r}'$ and $\mathbf{T}'$ are arbitrary, and $\boldsymbol{\Delta}^-$ is a g-inverse of (3.3).

To define $\boldsymbol{\Delta}^-$, we first let m be the number of values of y_j, $j = 1,\ldots,N$ that are non-zero. Furthermore, let $\mathbf{y}^-$ represent an $N \times 1$ vector with elements equal to $1/y_j$ if $y_j \neq 0$, and zero otherwise. Finally, let $\mathbf{i_y}$ represent an $N \times 1$ vector with elements equal to one if $y_j \neq 0$, and zero otherwise. With such definitions, we define a g-inverse of $\boldsymbol{\Delta}$ as

$$\boldsymbol{\Delta}^- = (N-1)\left[\mathbf{D_{y^-}}\mathbf{D_{y^-}} + \frac{1}{N-m}\mathbf{y}^-\mathbf{y}^{-\prime}\right]. \qquad (A.7)$$

Pre-multiplying this expression by $\boldsymbol{\Delta}$ yields $\boldsymbol{\Delta}\boldsymbol{\Delta}^- = \mathbf{D_{i_y}}$. Substituting this expression into (A.6),

$$\hat{\mathbf{C}} = \mathbf{L}\left(\mathbf{I}_N \otimes \frac{\mathbf{J}_{N\times n}}{n}\right) + \frac{1}{p}\mathbf{1}_p\mathbf{r}'_{1\times nN}[\mathbf{D_{y_0}} \otimes \mathbf{P}_n] + \mathbf{P}_p\mathbf{T}'_{p\times nN}(\mathbf{I}_N \otimes \mathbf{P}_n), \qquad (A.8)$$

where $\mathbf{D_{y_0}} = \mathbf{I}_N - \mathbf{D_{i_y}}$, a diagonal matrix with diagonal elements equal to zero for diagonal elements with $y_j \neq 0$, and one for diagonal elements with $y_j = 0$.

The general result given by (A.8) can be simplified by noting that for all $\mathbf{Y}_S$, $\mathbf{r}'_{1\times nN}[\mathbf{D_{y_0}} \otimes \mathbf{P}_n]\mathbf{Y}_S = 0$. Noting that the GMSE will not change with different choices of the arbitrary vector $\mathbf{r}$, eliminating the term that depends on $\mathbf{r}$ will not alter the predictor nor the GMSE, and simplifies the result. As a result, optimal estimators can be constructed using

$$\hat{\mathbf{C}} = \mathbf{L}\left(\mathbf{I}_N \otimes \frac{\mathbf{J}_{N\times n}}{n}\right) + \mathbf{P}_p\mathbf{T}'_{p\times nN}(\mathbf{I}_N \otimes \mathbf{P}_n). \qquad (A.9)$$

References

Basu, D., 1971. An essay on the logical foundations of survey sampling, Part 1. In: Godambe, V.P., Sprott, D.A. (Eds.), Foundations of Statistical Inference. Holt, Rinehart and Winston, Toronto, pp. 203–242.

Bellhouse, D.R., Rao, J.N.K., 1986. On the efficiency of prediction estimators in two-stage sampling. J. Statist. Plann. Inference 13, 269–281.

Bolfarine, H., Zacks, S., 1992. Prediction Theory for Finite Populations. Springer, New York.

Brewer, K.R.W., 1999. Design-based or prediction-based inference? Stratified random vs stratified balanced sampling. Int. Statist. Rev. 67, 35–47.

Brewer, K.R.W., 2002. Combined Survey Sampling Inference. Oxford University Press, New York.

Brewer, K.R.W., Hanif, M., Tam, S.M., 1988. How nearly can model-based prediction and design-based estimation be reconciled? J. Amer. Statist. Assoc. 83, 128–132.

Cassel, C.M., Särndal, C.E., Wretman, J.H., 1977. Foundations of Inference in Survey Sampling. Wiley, New York.

Fuller, W.A., Battese, G.E., 1973. Transformations for estimation of linear models with nested-error structure. Amer. Statist. Assoc. 68, 626–632.

Godambe, V.P., 1955. A unified theory of sampling from finite populations. J. Roy. Statist. Soc. Ser. B 17, 269–278.

Godambe, V.P., Joshi, V.M., 1965. Admissibility and Bayes estimation in sampling from finite populations: I. Ann. Math. Statist. 26, 1707–1722.

Graybill, F.A., 1983. Matrices with Applications in Statistics. Wadsworth International Group, Belmont, CA.

Hartley, H.O., Rao, J.N.K., 1968. A new estimation theory for sample surveys. Biometrika 55, 547–557.

Hartley, H.O., Rao, J.N.K., 1969. A new estimation theory for sample surveys, II. In: Godambe, V.P., Sprott, D.A. (Eds.), New Developments in Survey Sampling. Wiley, New York, pp. 147–169.

Hedayat, A., Sinha, B.K., 1991. Design and Inference in Finite Population Sampling. Wiley, New York.

Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling without replacement from finite population. J. Amer. Statist. Assoc. 47, 668–670.

Kempthorne, O., 1952. Design and Analysis of Experiments. Wiley, New York.

Lencina, V.B., 2002. Modelos de Efeitos Aleatórios E Populações Finitas. Ph.D. Dissertation. Department of Statistics, University of São Paulo, São Paulo, Brazil.

Li, W., 2003. Use of Random Permutation Models in Rate Estimation and Standardization. Ph.D Dissertation, Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst, MA.

Mukhopadhyay, P., 1984. Optimum estimation of a finite population variance under generalized random permutation models. Calcutta Statist. Assoc. Bull. 33, 93–104.

Mukhopadhyay, P., 2001. Topics in survey sampling. In: Lecture Notes in Statistics, Vol. 153. Springer, New York.

Neyman, J., 1934. On the two different aspects of the representation method: the method of stratified sampling and the method of purposive selection. J. Roy. Statist. Soc. 97, 558–625.

Neyman, J.K., Iwaszkiewicz et al., 1935. Statistical problems in agricultural experimentation, Roy. Statist. Soc. 2(Suppl.) 109–180.

Padmawar, V.R., Mukhopadhyay, P., 1985. Estimation under two-stage random permutation models. Metrika 25, 149–154.

Rao, J.N.K., 1997. Developments in sample survey theory: an appraisal. Canad. J. Statist. 25, 1–21.

Rao, J.N.K., 1999a. Some current trends in sample survey theory and methods. Sankhyā B 61, 1–25.

Rao, J.N.K., 1999b. Some recent advances in model-based small area estimation. Surv. Meth. 25, 175–186.

Rao, J.N.K., Bellhouse, D.R., 1978. Optimal estimation of a finite population mean under generalized random permutation models. J. Statist. Plann. Inference 2, 125–141.

Rao, T.J., 1984. Some aspects of random permutation models in finite population sampling theory. Metrika 31, 25–32.

Royall, R.M., 1969. An old approach to finite population sampling theory. J. Amer. Statist. Assoc. 63, 1269–1279.

Royall, R.M., 1976. The linear least-squares prediction approach to two-stage sampling. J. Amer. Statist. Assoc. 71, 657–664.

Särndal, C.-E., Swensson, B., Wretman, J., 1992. Model Assisted Survey Sampling. Springer, New York.

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

Searle, S.R., 1982. Matrix Algebra Useful for Statistics. Wiley, New York.

Stanek, III E.J., Singer, J.M., 2002a. Predicting realized random effects with clustered samples from finite populations with response error. Unpublished: (http://www-unix.oit.umass.edu/∼cluster/ed/outline/yr2002/c02ed31v4.pdf).

Stanek, III E.J., Singer, J.M., 2002b. Predicting realized cluster parameters from two stage sample of unequal size clustered populations. Unpublished: (http://www-unix.oit.umass.edu/∼cluster/ed/outline/yr2002/c02ed43v1.pdf).

Thompson, M.E., 1997. Theory of Sample Surveys. Chapman & Hall, London.

Valliant, R., Dorfman, A.H., Royall, R.M., 2000. Finite Population Sampling and Inference: A Prediction Approach. Wiley, New York.