Estimation in Multiple-Frame Surveys

Sharon LOHR and J. N. K. RAO

Multiple-frame surveys are commonly used to decrease costs of sampling or to reduce undercoverage that could occur if only one sampling frame were used. We describe potential uses and examples of multiple-frame surveys. We then derive optimal linear estimators and pseudo–maximum likelihood estimators for the population total when samples are taken independently from each frame using probability sampling designs. We explore the properties of these estimators theoretically and through a simulation study. We also derive variance estimators and discuss some practical problems that may be encountered in multiple-frame surveys.

KEY WORDS: Complex survey; Dual-frame survey; Population size; Pseudo–maximum likelihood estimation; Sampling weight.

1. INTRODUCTION

In classical finite-population sampling, a sample is selected using a probability sampling design from a single sampling frame containing all of the units in the target population. In many situations, however, it may be more practical to take samples from Q different sampling frames whose union covers the population of interest, and then combine the information from the samples to estimate population quantities.

In a multiple-frame survey, probability samples are drawn independently from the frames $A_1, \ldots, A_Q$, $Q \geq 2$. The union of the Q frames is assumed to cover the finite population of interest, $\mathcal{U}$. The frames may overlap, resulting in a possible $2^Q - 1$ nonoverlapping domains. When $Q = 2$, the survey is called a dual-frame survey.

Hartley (1962, 1974) showed that dual-frame surveys can cost far less than a single-frame survey that achieves the same precision. His applications concentrated on the situation where one frame is complete but expensive to sample; other frames are inexpensive to sample but incomplete. In many agricultural surveys, an area frame consists of segments of land; enumerators visit a probability sample of the segments. A list frame consists of the names and addresses of agricultural operators. The area frame is complete and insensitive to changes in farm ownership and activity, but very expensive to sample because of the in-person visits. The list frames are usually less costly to sample, particularly if the commodity of interest is concentrated in the operators on the list, but the lists may not include all producers of the commodity. The 1998 and 1999 U.S. equine estimates used three frames: the frame for the 1997 Census of Agriculture, an area frame, and a list frame of commercial equine operators (National Agricultural Statistics Service 1999).

Similar efficiencies can result from sampling rare populations (Kalton and Anderson 1986). For a survey studying characteristics of persons with AIDS, one sample could be taken from the frame for a general population health survey, whereas independent samples could be taken of clients of sexually transmitted disease clinics, drug treatment centers, and hospitals. The multiple-frame approach would be expected to produce more accurate estimates of characteristics of persons with human immunodeficiency virus (HIV), because the last three frames would provide larger numbers of HIV-positive individuals for the sample.

Hartley (1962, 1974) also discussed the use of multiple-frame surveys when all frames are incomplete. Iachan and Dennis (1993) described a multiple-frame survey of the homeless population in which the frames were homeless shelters, soup kitchens, and street areas. Not all homeless persons visit shelters, so including more frames likely reduces undercoverage bias. To improve response rates, different modes of interviewing may be used for the different frames. Haines and Pollock (1998) used multiple-frame surveys to estimate sizes of animal populations. Schoorl et al. (2000, pp. 146–147) described the design of a multiple-frame survey used for studying characteristics of Egyptian and Ghanaian immigrants in Italy, in which the sampling frames were mosques, health care centers, and other meeting places known to be frequented by the immigrants.

Multiple-frame surveys are becoming more common as new demands are placed on data from surveys. Rao (2003, p. 23) discussed the uses of multiple-frame surveys for small-area estimation, where a sample from an area frame may be supplemented by less-expensive samples from list frames. Madans, Ezzati-Rice, Cynamon, and Blumberg (2001) discussed multiple-frame surveys in the context of supplementing information from the U.S. National Health Interview Survey (NHIS). Additional surveys may be taken from different states and combined with information from NHIS for improved estimation at the state level.

In these applications, population units may be included in more than one frame. This differs from the *screening* multiple frame surveys described by González-Villalobos and Wallace (1996), in which the sampling frames are prescreened to remove overlap so that $A_1, \ldots, A_Q$ are disjoint. Population totals are easily estimated in screening multiple-frame surveys by summing the estimated population totals from each frame.

For dual-frame surveys, a number of estimators have been proposed for the population total $Y = \sum_{i=1}^{N} y_i$, where y_i is a quantity associated with population unit i and N is the number of population units in $\mathcal{U}$. The estimators of Hartley (1962, 1974) and Fuller and Burmeister (1972) minimize the variance among the class of linear unbiased estimators of Y; these estimators have minimum variance for a single response but use a different set of weights for each response variable. Bankier (1986) and Kalton and Anderson (1986) developed single-frame estimators in which observations are weighted according to their

Sharon Lohr is Thompson Industries Dean's Distinguished Professor of Statistics, Department of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287 (E-mail: *sharon.lohr@asu.edu*). J. N. K. Rao is Distinguished Research Professor, School of Mathematics and Statistics, Carleton University, Ottawa ON K1S 5B6, Canada (E-mail: *jrao@math.carleton.ca*).
This research was supported in part by the National Science Foundation (grant 0105852) and by a grant from the Natural Sciences and Engineering Research Council of Canada. The authors thank the associate editor and referees for their insightful and helpful comments, and thank Cherie Alf for collecting the data used in Section 9.

Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Theory and Methods
DOI 10.1198/016214506000000195

inclusion probabilities for the two frames. Skinner (1991) proposed raking ratio estimators for the situation when simple random samples are taken from each frame, and Skinner and Rao (1996) derived a pseudo–maximum likelihood estimator (PMLE) for Y for dual-frame surveys using complex designs. The single-frame estimator and PMLEs, unlike the Hartley and Fuller–Burmeister estimators, use the same set of weights for all response variables. Lohr and Rao (2000) compared the asymptotic efficiencies of these estimators in dual-frame surveys and found that the PMLE combined high efficiency with applicability to complex surveys. However, the PMLE was developed by Skinner and Rao (1996) only for dual-frame surveys; estimation methods for multiple-frame surveys using more than two frames are needed as such surveys become more prevalent.

In this article we develop optimal estimators and PMLEs for the multiple-frame setting with two or more frames. In Section 2 we define notation for the general estimation problem. In Sections 3 and 4 we derive optimal linear estimators and PMLEs of the population total for multiple-frame surveys. We generalize the single-frame approach to more than two frames in Section 5. In Section 6 we compare the estimators theoretically, and in Section 7 we compare them through a simulation study. We discuss variance estimation in Section 8, apply the estimators to a problem of estimating the total number of persons in three statistical societies in Section 9, and present some conclusions in Section 10.

2. ESTIMATION PROBLEM

Let $\mathcal{F}$ denote the index set of frames, $\mathcal{F} = \{1, 2, \dots, Q\}$. We assume that the union of the frames covers the population of interest. With Q frames, there are a possible $2^Q - 1$ distinct domains, defined by the subsets of $\mathcal{F}$. For $K \subseteq \mathcal{F}$, the domain defined by K is

$$D_K = \left(\bigcap_{i \in K} A_i\right) \cap \left(\bigcap_{j \notin K} A_j^c\right),$$

where c denotes complementation. The number of frames contributing to domain K is denoted by $|K|$. Let $N^{(q)}$ be the number of population units in frame A_q, and let N_K be the number of population units in domain K. Because the domains do not overlap, we have $\sum_{K \subseteq \mathcal{F}} N_K = N$, the population size. Although the frame sizes [$N^{(q)}$'s] are often known from other sources—$N^{(1)}$ might be, for example, the number of agricultural holdings in the list frame—the domain sizes N_K are often unknown and may be of independent interest.

Two three-frame surveys are shown in Figure 1. Figure 1(a) displays a three-frame design in which each frame is incomplete: Domains $\{1\}$, $\{2\}$, and $\{3\}$ appear only in frames A_1, A_2, and A_3 respectively; domains $\{1, 2\}$, $\{1, 3\}$, and $\{2, 3\}$ appear in exactly two frames; and domain $\{1, 2, 3\}$ is the intersection of the three frames. Figure 1(b) is also a three-frame design but has four domains, because frame A_1 is complete and frames A_2 and A_3 are incomplete but overlapping. Because frame A_1 is complete, a "1" appears in each domain name.

The population total Y may be written as the sum of the population totals in the distinct domains,

$$Y = \sum_{K \subseteq \mathcal{F}} Y_K, \tag{1}$$

where $Y_K = \sum_{i=1}^N \delta_i(K) y_i$ and $\delta_i(K) = 1$ if $i \in D_K$ and 0 otherwise. The number of population units in domain D_K is a special case of Y_K when $y_i = 1$ for all i: $N_K = \sum_{i=1}^N \delta_i(K)$. The problem of estimating the population total Y reduces to estimating each domain component Y_K, using the $|K|$ estimates from the different sampling frames.

Let $\mathcal{S}_q$ denote the probability sample from frame A_q, $q = 1, \dots, Q$. Because the samples from the Q frames are selected independently, the multiple-frame survey has $|K|$ independent estimators of Y_K. For $q \in K$, we estimate Y_K using the sample from frame A_q by

$$\hat{Y}_K^{(q)} = \sum_{i \in \mathcal{S}_q} w_i^{(q)} \delta_i(K) y_i. \tag{2}$$

The weights $w_i^{(q)}$ are the appropriate weights used for estimation from frame A_q; these may be the sampling weights (inverses of the probabilities of selection) or Hájek-type weights, where each sampling weight in frame A_q is multiplied by $N^{(q)}$/(sum of sampling weights) so that the new weights sum to $N^{(q)}$. Define the estimators of domain sizes N_K similarly using

$$\hat{N}_K^{(q)} = \sum_{i \in \mathcal{S}_q} w_i^{(q)} \delta_i(K). \tag{3}$$

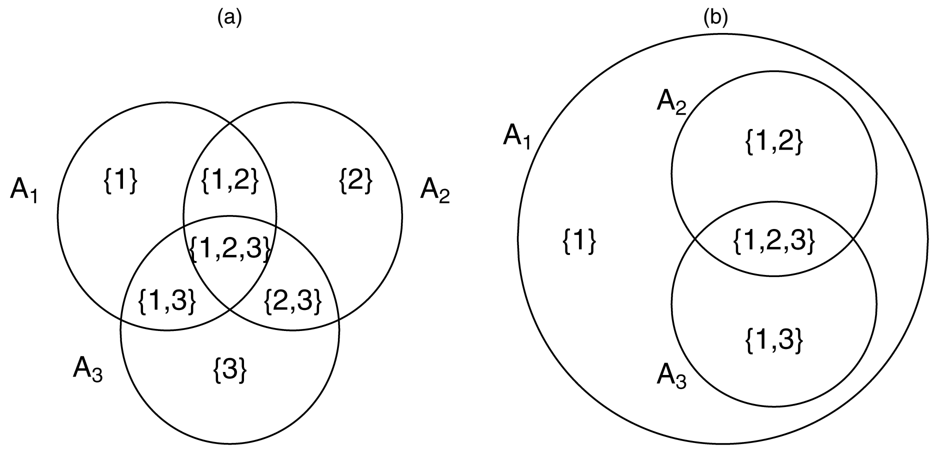

Figure 1. Two Three-Frame Survey Designs.

We assume throughout that we can correctly identify the domain membership of each unit in each of the surveys. This is a strong assumption that may not always be true in practice; we discuss it further in Sections 7 and 10. We do not need to be able to identify whether units are sampled from more than one frame, however.

3. OPTIMAL LINEAR ESTIMATORS OF THE POPULATION TOTAL

We first derive optimal estimators for general survey designs. Label the nonempty domains as $K_1, \dots, K_d$. The nonempty domains may be identifiable from the frame structure, as in Figure 1, or from the samples assuming that the overall sample size is sufficiently large to ensure that the probability of getting a zero sample size from a nonempty domain is negligible. Let $\boldsymbol{\theta}_q = (\theta_{K_1}^{(q)}, \dots, \theta_{K_d}^{(q)})^T$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_Q^T)^T$ be specified constants, and let $\hat{\mathbf{Y}}_q = (\hat{Y}_{K_1}^{(q)}, \dots, \hat{Y}_{K_d}^{(q)})^T$ and $\hat{\mathbf{N}}_q = (\hat{N}_{K_1}^{(q)}, \dots, \hat{N}_{K_d}^{(q)})^T$. Of course $\hat{Y}_{K_i}^{(q)} = \hat{N}_{K_i}^{(q)} = 0$ and $\theta_{K_i}^{(q)} = 0$ if $q \notin K_i$. (Note that $q \in K_i$ if domain K_i is part of frame A_q.) Define the $d \times d$ diagonal matrix $\boldsymbol{\Gamma}_q$ to have ith diagonal entry 1 if $q \in K_i$ and 0 otherwise, and let $\mathbf{1}_d$ be a d-vector of 1's. Then we can estimate Y by

$$\hat{Y}_{\boldsymbol{\theta}} = \sum_{q=1}^{Q} \boldsymbol{\theta}_q^T \boldsymbol{\Gamma}_q \hat{\mathbf{Y}}_q \tag{4}$$

for some value of $\boldsymbol{\theta}$. If a fixed vector $\boldsymbol{\theta}$ satisfies the constraint

$$\sum_{q=1}^{Q} \boldsymbol{\Gamma}_q \boldsymbol{\theta}_q = \mathbf{1}_d, \tag{5}$$

and if $E[\hat{Y}_K^{(q)}] = Y_K$ for each domain K and frame A_q, then $\hat{Y}_{\boldsymbol{\theta}}$ is an unbiased estimator of the population total Y. If the approximately unbiased Hájek estimator is used for $\hat{Y}_K^{(q)}$ in (4), then $\hat{Y}_{\boldsymbol{\theta}}$ is also approximately unbiased.

The generalized Hartley estimator uses the value $\boldsymbol{\theta}_H$ of $\boldsymbol{\theta}$ that minimizes the variance of $\hat{Y}_{\boldsymbol{\theta}}$ subject to the constraint in (5). Let $\mathbf{C}_q$ denote the covariance matrix of $\hat{\mathbf{Y}}_q$. Then

$$V(\hat{Y}_{\boldsymbol{\theta}}) = \sum_{q=1}^{Q} \boldsymbol{\theta}_q^T \boldsymbol{\Gamma}_q \mathbf{C}_q \boldsymbol{\Gamma}_q \boldsymbol{\theta}_q, \tag{6}$$

which is minimized when $\boldsymbol{\theta}_H$ is the solution to

$$\mathbf{A}\boldsymbol{\theta} = [\,\mathbf{0}_{dQ}^T \quad \mathbf{1}_d^T\,]^T, \tag{7}$$

where $\mathbf{A}$ is the $d(Q+1) \times dQ$ matrix with (i,j)th block given by

$$\mathbf{A}_{i,j} = \begin{cases} \boldsymbol{\Gamma}_i \left(\sum_{q=1}^{Q} \boldsymbol{\Gamma}_q\right)^{-1} \left(\mathbf{I} - \sum_{q=1}^{Q} \boldsymbol{\Gamma}_q\right) \mathbf{E}_i & \text{if } j = i \\ \boldsymbol{\Gamma}_i \left(\sum_{q=1}^{Q} \boldsymbol{\Gamma}_q\right)^{-1} \mathbf{E}_j & \text{if } i \neq j,\ i \leq Q \\ \boldsymbol{\Gamma}_j & \text{if } i = Q+1, \end{cases} \tag{8}$$

and $\mathbf{E}_i = \boldsymbol{\Gamma}_i \mathbf{C}_i \boldsymbol{\Gamma}_i$.

To generalize the Fuller–Burmeister estimator, we use the vectors of estimated domain sizes $\hat{\mathbf{N}}^{(q)}$. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_Q^T)^T$ be a $2dQ$-vector of parameters and define $\mathbf{G}_q = \mathrm{diag}(\boldsymbol{\Gamma}_q, \boldsymbol{\Gamma}_q)$. Then

$$\hat{Y}_{FB}(\boldsymbol{\beta}) = \sum_{q=1}^{Q} \boldsymbol{\beta}_q^T \mathbf{G}_q [\,\hat{\mathbf{Y}}_q^T \quad \hat{\mathbf{N}}_q^T\,]^T, \tag{9}$$

where $\boldsymbol{\beta}$ satisfies the constraints

$$\sum_{q=1}^{Q} \mathbf{G}_q \boldsymbol{\beta}_q = [\,\mathbf{1}_d^T \quad \mathbf{0}_d^T\,]^T.$$

Define the covariance matrix

$$\mathbf{D}_q = \mathrm{cov}\begin{bmatrix} \hat{\mathbf{Y}}_q \\ \hat{\mathbf{N}}_q \end{bmatrix} = \begin{bmatrix} \mathbf{C}_q & \mathbf{D}_{q12} \\ \mathbf{D}_{q12}^T & \mathbf{D}_{q22} \end{bmatrix}.$$

Then

$$V(\hat{Y}_{FB}) = \sum_{q=1}^{Q} \boldsymbol{\beta}_q^T \mathbf{G}_q \mathbf{D}_q \mathbf{G}_q \boldsymbol{\beta}_q. \tag{10}$$

Similarly to (7), it is shown that $V(\hat{Y}_{FB})$ is minimized subject to the constraints when $\boldsymbol{\beta}_{FB}$ satisfies

$$\mathbf{B}\boldsymbol{\beta} = [\,\mathbf{0}_{2dQ}^T \quad \mathbf{1}_d^T \quad \mathbf{0}_d^T\,]^T, \tag{11}$$

where the blocks of the $2d(Q+1) \times 2dQ$ matrix $\mathbf{B}$ are defined similarly to the blocks of $\mathbf{A}$ in (8) with $\boldsymbol{\Gamma}_i$ replaced by $\mathbf{G}_i$ and $\mathbf{E}_i$ replaced by $\mathbf{G}_i\mathbf{D}_i\mathbf{G}_i$.

For the dual-frame case ($Q = 2$) with domains $K_1 = \{1\}$, $K_2 = \{1, 2\}$, and $K_3 = \{2\}$, the optimal linear estimators may be written in simpler form, giving the Hartley and Fuller–Burmeister estimators. The solution to (7) is $\boldsymbol{\theta}_1 = [1, t, 0]^T$ and $\boldsymbol{\theta}_2 = [0, 1 - t, 1]^T$, where $t = (\mathbf{C}_2[2,2] + \mathbf{C}_2[2,3] - \mathbf{C}_1[1,2])/(\mathbf{C}_1[2,2] + \mathbf{C}_2[2,2])$. The value of t may be outside the interval $[0, 1]$, but this occurs only when one of the covariances $\mathbf{C}_2[2,3]$ or $\mathbf{C}_1[1,2]$ is very large. The solution to (11) is $\boldsymbol{\beta}_1 = [1, t_1, 0, 0, t_2, 0]^T$ and $\boldsymbol{\beta}_2 = [0, 1 - t_1, 1, 0, -t_2, 0]$, where

$$\begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = -\mathrm{cov}\begin{bmatrix} \hat{Y}_{\{1,2\}}^{(1)} - \hat{Y}_{\{1,2\}}^{(2)} \\ \hat{N}_{\{1,2\}}^{(1)} - \hat{N}_{\{1,2\}}^{(2)} \end{bmatrix}^{-1} \times \begin{bmatrix} \mathrm{cov}(\hat{Y}_{\{1\}}^{(1)} + \hat{Y}_{\{2\}}^{(2)} + \hat{Y}_{\{1,2\}}^{(2)}, \hat{Y}_{\{1,2\}}^{(1)} - \hat{Y}_{\{1,2\}}^{(2)}) \\ \mathrm{cov}(\hat{Y}_{\{1\}}^{(1)} + \hat{Y}_{\{2\}}^{(2)} + \hat{Y}_{\{1,2\}}^{(2)}, \hat{N}_{\{1,2\}}^{(1)} - \hat{N}_{\{1,2\}}^{(2)}) \end{bmatrix}.$$

Because each is derived by minimizing the variance, $\hat{Y}_H$ is optimal among all unbiased estimators linear in $(\hat{\mathbf{Y}}_1, \dots, \hat{\mathbf{Y}}_Q)$ and $\hat{Y}_{FB}$ is optimal among all unbiased estimators linear in $(\hat{\mathbf{Y}}_1, \dots, \hat{\mathbf{Y}}_Q)$ and $(\hat{\mathbf{N}}_1, \dots, \hat{\mathbf{N}}_Q)$. Both estimators of Y may be written in the form

$$\hat{Y} = \sum_{q=1}^{Q} \sum_{i \in \mathcal{S}_q} \tilde{w}_i^{(q)} y_i \tag{12}$$

for modified weights $\tilde{w}_i^{(q)}$. For example, for $\hat{Y}_H(\boldsymbol{\theta}_H)$, the modified weight for an observation in domain K_j is $\tilde{w}_{iH}^{(q)} = w_i^{(q)} \delta_i(K_j) \theta_{K_j}^{(q)}$, where the value of $\theta_{K_j}^{(q)}$ comes from the optimal vector $\boldsymbol{\theta}_H$ which is a function of the covariances. As a result, the weights $\tilde{w}_i^{(q)}$ change for different response variables y for both $\hat{Y}_H(\boldsymbol{\theta})$ and $\hat{Y}_{FB}(\boldsymbol{\beta})$. This can lead to survey estimates that are not internally consistent. For example, suppose that Y_1 = total nursing home costs for patients under age 85, Y_2 = total nursing home costs for patients age 85 and over, and Y_3 = total nursing home costs for all patients. In multiple-frame surveys with complex designs, it is generally the case that $\hat{Y}_1 + \hat{Y}_2 \neq \hat{Y}_3$ when the generalized Hartley or Fuller–Burmeister estimators are used.

4. PSEUDO–MAXIMUM LIKELIHOOD ESTIMATION

The PMLE proposed in this section uses the same set of weights for each response variable, and thus avoids the internal consistency problems of the optimal linear estimators derived in Section 3. In Section 4.1 we first extend Skinner's (1991) approximate MLE of the total Y, with simple random samples from dual frames, to multiple frames. We then extend the results to complex samples in Section 4.2, using the results in Section 4.1.

4.1 Approximate Maximum Likelihood Estimation With Simple Random Samples

Assume that a simple random sample of size $n^{(q)}$ is taken from frame A_q, for $q = 1, \dots, Q$. Let $\mathbf{f} = (f^{(1)}, \dots, f^{(Q)})^T$ be the Q-vector of sampling fractions with $f^{(q)} = n^{(q)}/N^{(q)}$, noting that for simple random sampling, $w_i^{(q)} = 1/f^{(q)}$. Then $n_K^{(q)} = \hat{N}_K^{(q)} f^{(q)}$ is the (random) sample size from frame A_q that is in domain K, where $\hat{N}_K^{(q)}$ is defined in (3). In the multiple-frame situation, we may write

$$Y = \sum_{K \subseteq \mathcal{F}} Y_K = \mathbf{N}^T \bar{\mathbf{Y}},$$

where $\mathbf{N} = (N_{K_1}, \dots, N_{K_d})^T$, $\mathbf{Y} = (Y_{K_1}, \dots, Y_{K_d})^T$, and $\bar{\mathbf{Y}}$ is the vector of corresponding domain means. We find approximate MLEs for $\bar{\mathbf{Y}}$ and $\mathbf{N}$, so that the approximate MLE for Y is

$$\hat{Y} = \hat{\mathbf{N}}^T \hat{\bar{\mathbf{Y}}} = \sum_{K \subseteq \mathcal{F}} \hat{N}_K \hat{\bar{Y}}_K. \qquad (13)$$

For simple random samples under appropriate regularity conditions, the central limit theorem applies to

$$\bar{y}_K^{(q)} = \sum_{i \in \mathcal{S}_q} \delta_i(K) y_i / n_K^{(q)}.$$

For sufficiently large sample sizes, $\bar{y}_K^{(q)} | n_K^{(q)}$ is approximately distributed as $N(\bar{Y}_K, \sigma_K^2 / n_K^{(q)})$.

For each q, under simple random sampling with small sampling fractions, $\{n_K^{(q)}\}$ approximately follows a multinomial distribution with sample size $n^{(q)}$ and success probabilities $N_K/N^{(q)}$ for domains K in frame A_q. Thus the likelihood of $(\bar{\mathbf{Y}}, \mathbf{N})$ may be approximated by $\mathcal{L}(\bar{\mathbf{Y}}, \mathbf{N}) = \mathcal{L}_1(\bar{\mathbf{Y}}) \mathcal{L}_2(\mathbf{N})$, where

$$\log \mathcal{L}_1(\bar{\mathbf{Y}}) = \text{constant} - \frac{1}{2} \log \sigma_K^2 - \sum_{q=1}^{Q} \sum_{K: q \in K} \frac{n_K^{(q)} (\bar{y}_K^{(q)} - \bar{Y}_K)^2}{2\sigma_K^2} \qquad (14)$$

and

$$\log \mathcal{L}_2(\mathbf{N}) = \text{constant} + \sum_{q=1}^{Q} \sum_{K: q \in K} n_K^{(q)} \log N_K = \text{constant} + \sum_{q=1}^{Q} \sum_{K: q \in K} \hat{N}_K^{(q)} f^{(q)} \log N_K, \qquad (15)$$

with

$$N^{(q)} = \sum_{K: q \in K} N_K, \qquad q = 1, \dots, Q. \qquad (16)$$

Maximizing (14) gives

$$\hat{\bar{Y}}_K = \frac{\sum_{q \in K} n_K^{(q)} \bar{y}_K^{(q)}}{\sum_{q \in K} n_K^{(q)}} = \frac{\sum_{q \in K} f^{(q)} \hat{Y}_K^{(q)}}{\sum_{q \in K} f^{(q)} \hat{N}_K^{(q)}}. \qquad (17)$$

To maximize (15), define $\mathbf{M}$ to be the $d \times Q$ matrix whose (i,j)th entry is 1 if $j \in K_i$ and 0 otherwise. Let $\hat{\mathbf{H}}$ be the $d \times Q$ matrix with

$$\hat{\mathbf{H}}_{iq} = \begin{cases} \hat{N}_{K_i}^{(q)} & \text{if } q \in K_i \\ 0 & \text{otherwise.} \end{cases}$$

For simple random sampling, $\hat{N}_K^{(q)} = (N^{(q)}/n^{(q)}) n_K^{(q)}$. Also define $\mathbf{h} = (N^{(1)}, N^{(2)}, \dots, N^{(Q)})^T$. The constraint in (16) means that the vector that maximizes (15), $\hat{\mathbf{N}}$, must satisfy $\mathbf{M}^T \hat{\mathbf{N}} = \mathbf{h}$. It is shown in the Appendix that $\hat{\mathbf{N}}$ is the solution to

$$\begin{bmatrix} (\mathbf{I} - \mathbf{M}\mathbf{M}^+)(\operatorname{diag} \hat{\mathbf{N}})^{-1} \hat{\mathbf{H}} \mathbf{f} \\ \mathbf{M}^T \hat{\mathbf{N}} - \mathbf{h} \end{bmatrix} = \mathbf{0}_{Q+d}, \qquad (18)$$

where $\mathbf{M}^+$ is the Moore–Penrose generalized inverse of $\mathbf{M}$. If $\mathbf{M}$ is nonsingular, then $\hat{\mathbf{N}} = \mathbf{M}^{-T} \mathbf{h}$; in that case we know the true values of $\mathbf{N}$.

The system of equations in (18) is not linear in $\hat{\mathbf{N}}$; its solution involves solving polynomials of degree $d - 1$. By using the implicit mapping theorem (see the App.), we find a linear approximation $\tilde{\mathbf{N}}$ to $\hat{\mathbf{N}}$; the linearized estimator $\tilde{\mathbf{N}}$ solves

$$\begin{bmatrix} (\mathbf{I} - \mathbf{M}\mathbf{M}^+)(\operatorname{diag} \mathbf{N})^{-1} (\operatorname{diag} \mathbf{M}\mathbf{f}) \\ \mathbf{M}^T \end{bmatrix} \tilde{\mathbf{N}} = \begin{bmatrix} (\mathbf{I} - \mathbf{M}\mathbf{M}^+)(\operatorname{diag} \mathbf{N})^{-1} \hat{\mathbf{H}} \mathbf{f} \\ \mathbf{h} \end{bmatrix}. \qquad (19)$$

Note that (19) may be used to find the asymptotic variance of $\hat{\mathbf{N}}$, which is the same as the asymptotic variance of $\tilde{\mathbf{N}}$.

Equation (19) also may be used iteratively to calculate $\hat{\mathbf{N}}$ in (18). Let $\tilde{\mathbf{N}}_0$ be a vector of initial estimates for $\mathbf{N}$ and define $\tilde{\mathbf{N}}_{k+1}$ by the solution to

$$\begin{bmatrix} (\mathbf{I} - \mathbf{M}\mathbf{M}^+)(\operatorname{diag} \tilde{\mathbf{N}}_k)^{-1} (\operatorname{diag} \mathbf{M}\mathbf{f}) \\ \mathbf{M}^T \end{bmatrix} \tilde{\mathbf{N}}_{k+1} = \begin{bmatrix} (\mathbf{I} - \mathbf{M}\mathbf{M}^+)(\operatorname{diag} \tilde{\mathbf{N}}_k)^{-1} \hat{\mathbf{H}} \mathbf{f} \\ \mathbf{h} \end{bmatrix}. \qquad (20)$$

If the initial estimate $\tilde{\mathbf{N}}_0$ is reasonably good, then this procedure will converge quickly to $\hat{\mathbf{N}}$. In practice, one could use either the estimate from the frame with the largest sample size in domain K or the average of the $\hat{N}_K^{(q)}$'s as an initial estimate of N_K.

In the dual-frame case ($Q = 2$), (18) gives the quadratic equation in $\hat{N}_{\{1,2\}}$ found in equation (5) of Skinner and Rao (1996). Solving (19) gives $\tilde{N}_{\{1,2\}} = [N_{\{2\}} n^{(1)} \hat{N}^{(1)}_{\{1,2\}} + N_{\{1\}} n^{(2)} \hat{N}^{(2)}_{\{1,2\}}]/[N_{\{2\}} n^{(1)} + N_{\{1\}} n^{(2)}]$.

4.2 Pseudo-Maximum Likelihood Estimators in Complex Surveys

Because it is based on the maximum likelihood principle, $\hat{Y}$ given in (13) is asymptotically efficient when a simple random sample is taken from each frame. With complex survey designs, the estimator in (13) is consistent but not necessarily asymptotically efficient. To adapt the estimator to complex survey designs, following Skinner and Rao (1996), we treat each $f^{(q)} = n^{(q)}/N^{(q)}$ in (17), (18), and (19) as a quantity to be chosen based on the sampling design for frame A_q. Let $\tilde{n}^{(q)}$ be the actual sample size taken from frame A_q and let $d^{(q)}(y) = V(\hat{Y}^{(q)}_{A_q})/V_{\text{SRS}}(\hat{Y}^{(q)}_{A_q})$ be the design effect for estimating Y_{A_q} using sample $\mathcal{S}_q$, where V_{SRS} denotes the variance under simple random sampling. If all variables of interest y had approximately the same design effect $d^{(q)}$, then we would set $n^{(q)}$ to be the effective sample size, $n^{(q)} = \tilde{n}^{(q)}/d^{(q)}$ and $f^{(q)} = n^{(q)}/N^{(q)}$.

In practice, however, variables have different design effects; choosing the design effect for one variable may lead to suboptimal estimates for other variables. In the dual-frame case, Skinner and Rao (1996) proposed using the design effects $d^{(q)}$ from using $\mathcal{S}_q$ to estimate the population size of the intersection domain, $N_{\{1,2\}}$. This does not extend well to situations with more than two frames and multiple intersection domains. A simple alternative, following Chu, Brick, and Kalton (1999), is to use a compromise value of the design effect that works well for the most important variables by, for example, averaging the design effects for the main variables.

For complex survey designs, $\hat{Y}_{\text{PML}}$ is defined by (13), (17), and (18), where we use the weighted estimators $\hat{Y}^{(q)}_K$ in (2) and $\hat{N}^{(q)}_K$ in (3) and use design effects in setting the values of $f^{(q)}$. For any of the choices for **f** discussed earlier, $\hat{Y}_{\text{PML}}$ is a consistent estimator for Y. If the design effects chosen for **f** are far from those for a variable of interest, then there may be some loss of efficiency relative to the optimal Fuller–Burmeister estimator, but the PMLE still will be approximately unbiased.

The PMLE has the practical advantage of using the same set of weights for all variables. We may write the PMLE of a population total in the form of (12). Here, we set

$$\tilde{w}^{(q)}_{i,\text{PML}} = w^{(q)}_i f^{(q)} \sum_{K:\, q\in K} \frac{\hat{N}_K \delta_i(K)}{\sum_{j\in K} f^{(j)} \hat{N}^{(j)}_K}. \tag{21}$$

The weights $\tilde{w}^{(q)}_{i,\text{PML}}$ ensure internal consistency, unlike the Hartley and Fuller–Burmeister weights.

The PMLE, as presented here, assumes that the frame sizes $N^{(q)}$ are known; the domain sizes N_K may or may not be known. It may be modified for the unknown frame size case by replacing $N^{(q)}$ by an estimator $\hat{N}^{(q)} = \sum_{i\in\mathcal{S}_q} w^{(q)}_i$ and including the variability of the $\hat{N}^{(q)}$'s in the variance estimation (see Sec. 8).

4.3 Collapsing Domains

The theory in Sections 4.1 and 4.2 depends on large-sample approximations. For large samples, the distribution of $\bar{y}^{(q)}_K$ given $n^{(q)}_K$ is approximately normal. In small samples, however, $n^{(q)}_K$ may be 0, in which case $\bar{y}^{(q)}_K$ is not defined. Even if the sample sizes in all domains are positive, some of the domains may be small, and thus the PMLEs based on estimating the domain sizes may be unstable. In this section we provide an alternate method for estimation with multiple-frame surveys for this situation.

We illustrate the method for the design in Figure 1(a). We first combine frames A_1 and A_2 using dual-frame methods. Let $A_4 = A_1 \cup A_2$. The weight for an observation originally in frame A_q ($q = 1, 2$) for the "new" frame A_4 is $w^{(4)}_i = w^{(q)}_i f^{(q)} \sum_{K:\, q\in K} \hat{N}_K \delta_i(K)/[\sum_{j\in K} f^{(j)} \hat{N}^{(j)}_K]$, where K can be any of the three domains $A_1 \cap A_2$, $A_1 \cap A_2^c$, or $A_1^c \cap A_2$. Then $\hat{N}^{(4)} = N^{(1)} + N^{(2)} - \hat{N}_{A_1\cap A_2}$. Now combine frames A_3 and A_4 using dual-frame methods, treating $\hat{N}^{(4)}$ as though it were the true frame population size. The final weights are $\tilde{w}^q_i = w^{(q)}_i f^{(q)} \sum_{K:\, q\in K} \hat{N}_K \delta_i(K)/[\sum_{j\in K} f^{(j)} \hat{N}^{(j)}_K]$ for $q = 3, 4$. The final weights for this method depend on the order in which the frames are combined. We recommend starting with the pair of frames for which the estimated relative variance of $\hat{N}_{\{i,j\}}$ is the smallest.

5. SINGLE–FRAME ESTIMATION METHODS

The estimators in Sections 3 and 4 use separate sets of estimators from the Q frames, and then combine them to estimate Y. Single-frame methods, proposed by Bankier (1986) and Kalton and Anderson (1986) in the dual-frame setting, combine the observations into a single frame and then adjust the weights in the intersection domain to reflect the possibility that units in the intersection domain may be selected from each sample. In this section we discuss the use of single-frame methods with more than two frames.

Bankier's (1986) estimator requires identifying which units from the different samples are distinct. Such identification could be cumbersome with multiple frames. The Kalton and Anderson (1986) single-frame estimator does not share this difficulty. To extend the estimator to multiple frames, let

$$\tilde{w}^{(q)}_{i,S} = \sum_{K\subseteq\mathcal{F}} \delta_i(K) \left(\sum_{j\in K} \frac{1}{w^{(j)}_i} \right)^{-1}. \tag{22}$$

Then the single-frame estimator is

$$\hat{Y}_S = \sum_{q=1}^{Q} \sum_{i\in\mathcal{S}_q} \tilde{w}^{(q)}_{i,S} y_i. \tag{23}$$

Note that

$$E[\hat{Y}_S] = \sum_{q=1}^{Q} E\left[\sum_{i\in\mathcal{S}_q} \sum_{K\subseteq\mathcal{F}} \delta_i(K) \left(\sum_{j\in K} \frac{1}{w^{(j)}_i} \right)^{-1} y_i \right]$$
$$= \sum_{K\subseteq\mathcal{F}} \sum_{i\in\mathcal{U}_K} \sum_{q\in K} \delta_i(K) \pi^{(q)}_i y_i \left(\sum_{j\in K} \frac{1}{w^{(j)}_i} \right)^{-1},$$

where $\pi_i^{(q)} = P\{i \in \mathcal{S}_q\}$ and $\mathcal{U}_K$ is the set of population units in domain K. If $w_i^{(j)} = 1/\pi_i^{(j)}$ for each j, then $\hat{Y}_S$ is unbiased for Y, that is, $E[\hat{Y}_S] = Y$.

If each sample is self-weighted (i.e., $w_i^{(q)} = w_j^{(q)} = w^{(q)}$ for all i, j, and q), then $\hat{Y}_{SF}$ can be written in the same form as (4). In that case, (2), (22), and (23) imply that

$$\hat{Y}_S = \sum_{q=1}^{Q} \sum_{K:q\in K} \left(w^{(q)} \sum_{j\in K} \frac{1}{w^{(j)}} \right)^{-1} \sum_{i\in\mathcal{S}_q} w^{(q)} \delta_i(K) y_i$$

$$= \sum_{q=1}^{Q} \boldsymbol{\theta}_{q,S}^T \boldsymbol{\Gamma}_q \hat{\mathbf{Y}}_q,$$

with

$$\boldsymbol{\theta}_{q,S}^T = \left[\left(w^{(q)} \sum_{j\in K_1} 1/w^{(j)} \right)^{-1}, \ldots, \left(w^{(q)} \sum_{j\in K_d} 1/w^{(j)} \right)^{-1} \right].$$

The single-frame estimator may be adjusted to known frame sizes $N^{(q)}$ using either raking ratio estimation (Bankier 1986; Rao and Skinner 1996) or regression estimation. The weights are easily raked by iteratively setting the new weight for observation i in frame A_q equal to $\tilde{w}_{i,S}^{(q)}(N^{(q)}/\sum_j \tilde{w}_{j,S}^{(q)})$ for each frame in turn.

The single-frame estimator is easy to calculate for any number of frames when all of the separate surveys are self-weighted. But because $\boldsymbol{\theta}$ does not have to be the optimal Hartley value, we have that $V(\hat{Y}_S) \geq V(\hat{Y}_H)$. If the surveys are not self-weighted, then the single-frame estimator is more problematic, because it requires knowing what the weight of an observation would be in each of the frames, not just the survey in which the observation was selected. Single-frame estimators rely only on the sampling weights; unlike the other methods, they do not require calculation of variances or design effects. The raking ratio single-frame estimator also requires knowledge or external estimates of the frame sizes $N^{(q)}$.

O'Muircheartaigh and Pedlow (2002) studied the effects of combining two samples for the National Longitudinal Survey of Youth. They argued that for their study and weights, the single-frame approach produced a greater effective sample size than a Hartley-type optimal estimator; however, the optimal estimator that they used was not adjusted for true design effects but was adjusted only for the effect of differential weighting on the variability. They also noted that in many cases, calculating the inclusion probabilities for the other frames would be difficult.

Skinner, Holmes, and Holt (1997) used single-frame estimation in the context of multivariate stratification. Independent subsamples were selected by stratified random sampling with respect to each stratification variable related to the corresponding study variables. The data from the separate subsamples were then pooled using single-frame estimation.

6. ASYMPTOTIC VARIANCE OF ESTIMATORS

The asymptotic variance of the estimators considered in Sections 3–5 may be expressed in the form

$$V(\hat{Y}) = \sum_{q=1}^{Q} \boldsymbol{\eta}_q^T \mathbf{G}_q \mathbf{D}_q \mathbf{G}_q \boldsymbol{\eta}_q.$$

For the generalized Hartley estimator, $\boldsymbol{\eta}_{q,H}$ is taken to be $[\boldsymbol{\theta}_{q,H}^T, \ \mathbf{0}]^T$; for the generalized Fuller–Burmeister estimator, $\boldsymbol{\eta}_{q,FB} = \boldsymbol{\beta}_{q,FB}$. For the single-frame method, with each sample self-weighted, we have $\boldsymbol{\eta}_{q,S}^T = (\theta_{q,S}^T, 0^T)$, where $\theta_{q,S} = f^{(q)}(\operatorname{diag}\mathbf{Mf})^{-1}\mathbf{1}_d$.

Equations (13), (17), and (19) may be used to find the asymptotic variance of $\hat{Y}_{\text{PML}}$. Using the delta method and the fact that $\mathbf{MM}^+$ is the projection matrix onto the column space of $\mathbf{M}$, the asymptotic variance of $\hat{Y}_{\text{PML}}$ has

$$\boldsymbol{\eta}_{q,\text{PML}}^T = f_{\text{PML}}^{(q)} \big[\mathbf{1}_d^T (\operatorname{diag}\mathbf{Mf}_{\text{PML}})^{-1}, \\ - \bar{\mathbf{Y}}^T \mathbf{P}^{-1} \mathbf{MM}^T (\operatorname{diag}\mathbf{Mf}_{\text{PML}})^{-1} \big],$$

with

$$\mathbf{P} = (\operatorname{diag}\mathbf{N})^{-1} (\operatorname{diag}\mathbf{Mf}_{\text{PML}}) (\mathbf{I} - \mathbf{MM}^+) \\ \times (\operatorname{diag}\mathbf{Mf}_{\text{PML}}) (\operatorname{diag}\mathbf{N})^{-1} + \mathbf{MM}^T.$$

The qth component of $\mathbf{f}_{\text{PML}}$ is $f_{\text{PML}}^{(q)} = n^{(q)}/N^{(q)}$, where $n^{(q)}$ is the effective sample size for frame A_q.

The asymptotic variance of the modified PML method in Section 4.3, based on collapsing domains, has a more complex form because it requires iterating the PML method for subsets of two frames. Relevant formulas may be obtained from the authors.

The asymptotic variances of the estimators may be used to construct optimal designs for multiple-frame surveys. Because the asymptotic variance of $\hat{Y}$ depends on the domain mean and covariance structure for each response, it may be more practical to find an optimal design for estimating the population size N, using anticipated design effects from the three surveys. For the PML method, this becomes a problem in setting $\mathbf{f}_{\text{PML}}$ to minimize the anticipated variance.

7. SIMULATION STUDY

We performed a simulation study to explore the finite-sample properties of the estimators in Sections 3–5. Finite populations were generated from the three-frame setup in Figure 1(a) using a random-effects model with groups of size five inside each domain. For domain K, observations y_{ij} were generated from the model

$$y_{ij} = \mu_K + \alpha_i + \epsilon_{ij}, \qquad i = 1, \ldots, N_K/5, \ j = 1, \ldots, 5,$$

where $\alpha_i \sim \text{N}(0, 1)$ and $\epsilon_{ij} \sim \text{N}(0, 1)$. We also studied the three-frame setup in Figure 1(b) and two dual-frame designs through simulation, but the general results were similar, and we do not report them here.

We used a factorial design with fractional factorials used in factors 4 and 5 for the simulation study, with the following factors:

1. Sample sizes $\tilde{n}^{(1)}$, $n^{(2)}$, and $n^{(3)}$ [100 or 200]
2. Cluster sample from frame A_1 [yes or no]
3. Misclassification [yes or no]
4. Domain population sizes $\mathbf{N}^T = (N_{\{1\}}, N_{\{1,2\}}, N_{\{1,3\}}, N_{\{1,2,3\}}, N_{\{2\}}, N_{\{2,3\}}, N_{\{3\}})$ [each was 10,000 or 30,000]
5. Mean vector for the population, $\boldsymbol{\mu}^T = (\mu_{\{1\}}, \mu_{\{1,2\}}, \mu_{\{1,3\}}, \mu_{\{1,2,3\}}, \mu_{\{2\}}, \mu_{\{2,3\}}, \mu_{\{3\}})$ [(0, 0, 0, 0, 0, 0, 0), (5, 4, 3, 3, 2, 2, 1), (−10, 0, 0, 0, 0, 0, 0), (−10, 0, 0, 0, 10, 10, 5), (5, 5, 5, 5, 0, 0, 0), or (5, 5, 5, 5, 10, 10, 5)].

The factor levels were chosen to illustrate the effects of disparate sampling fractions and mean vectors on the estimators, with emphasis on the cases when either frame A_1 or domain $\{1\}$ have different means than the other domains. If a cluster sample was drawn from frame A_1, then $\tilde{n}^{(1)}/5$ of the groups used to generate the population were sampled. Simple random samples were drawn from frames A_2 and A_3 (and from frame A_1 if a cluster sample was not drawn).

To study the effects of misclassification, we also altered the domain classification of approximately 9.75% of the observations selected for the samples from each frame. For each observation i from the frame A_1 sample, two independent Bernoulli(.05) observations, u_{i2} and u_{i3}, were generated; if $u_{ij} = 1$, then the frame A_j membership for observation i was changed. Thus an observation from the frame A_1 sample in domain $\{1, 3\}$ remained in that domain if $u_{i2} = u_{i3} = 0$, was misclassified to domain $\{1, 2, 3\}$ if $u_{i2} = 1$ and $u_{i3} = 0$, was misclassified to domain $\{1\}$ if $u_{i2} = 0$ and $u_{i3} = 1$, and was misclassified to domain $\{1, 2\}$ if $u_{i2} = u_{i3} = 1$. The same procedure was followed to misclassify observations in frames A_2 and A_3.

In implementing the methods, we used the singular value decomposition to solve the linear equations (7) and (11) used in the Hartley and Fuller–Burmeister estimators. For the PMLEs, we solved (19) iteratively to estimate the domain population sizes. When a cluster sample was taken in frame A_1, we used the effective sample size $n^{(1)} = \tilde{n}^{(1)}/d$ in (18) and (13); we set d as the average of the design effects for the domain sizes in frame A_1.

All computations were performed in version 1.8.1 of R (www.r-project.org) on a 1.4-GHz PC running the Linux operating system. A total of 2,000 runs were done for each setting; the square root of the empirical mean squared error (EMSE) for each estimator for a subset of the simulation runs is reported in Tables 1–4. Eight estimators are compared. The Ave method simply averages estimators in the overlap domains; it is of the form $\hat{Y}_{\boldsymbol{\theta}}$ in (4) with $\theta^{(q)}_{K_i} = 1/(\text{number of frames in domain } K_i)$ for $q \in K_i$. Hart and FB are the Hartley and Fuller–Burmeister methods. PML refers to the full PML method, PML12 uses the method of combining frames in Section 4.3 with frames A_1 and A_2 combined first, and PML23 combines frames with A_2 and A_3 combined first. SF refers to the unadjusted single-frame method from Section 5, and SFrake is the single-frame method raked to the frame sizes $N^{(q)}$.

As can be seen in Tables 1–4, in almost every simulation experiment the PML method had the smallest or close to the smallest EMSE for estimating N and Y. When the domain means were equal, as in Table 4, the eight methods studied performed similarly, with the FB method having a slightly larger EMSE than the other methods; in this case suboptimal weighting of the domains had less effect on the estimated total, $\hat{Y}$. When the domain means were unequal, as in Tables 1–3, the PML method had superior performance.

In particular, the EMSE from the PML method was almost always smaller than the EMSE from the FB method. The FB method is theoretically optimal, but in practice, for the setup in Figure 1(a), an 8×8 covariance matrix must be estimated for each of the three samples, and the extra variability in estimating the covariances leads to a larger EMSE for the FB method. Although the Hartley method performed well with small samples when $Q = 2$, it performed relatively poorly in the three-frame setting. In some iterations, one or more components of $\hat{\boldsymbol{\theta}}_H$ were negative, leading to domain estimates such as $\hat{Y}_{\{1,2,3\}} = -.5\hat{Y}^{(1)}_{\{1,2,3\}} + 1.2\hat{Y}^{(2)}_{\{1,2,3\}} + .3\hat{Y}^{(1)}_{\{1,2,3\}}$ rather than a convex combination. The instability of the three-frame Hartley

Table 1. Simulation Results for the Population With $\mathbf{N}^T = (10{,}000, 10{,}000, 10{,}000, 10{,}000, 10{,}000, 10{,}000, 10{,}000)$ and $\boldsymbol{\mu}^T = (-10, 0, 0, 0, 10, 10, 5)$

$\tilde{n}^{(1)}$	Clus?	$n^{(2)}$	$n^{(3)}$	Est.	Ave	Hart	FB	PML	PML12	PML23	SF	SFrake
100	No	100	100	N	**1,698**	**1,748**	**1,748**	**1,696**	**1,757**	**1,773**	**1,698**	**1,701**
				Y	27,083	27,607	12,050	**11,445**	18,469	13,172	27,083	**11,677**
100	Yes	100	100	N	2,628	2,391	2,391	**2,141**	2,343	2,416	2,628	2,687
				Y	44,767	50,779	17,425	**14,903**	35,279	**14,624**	44,767	17,784
200	No	100	100	N	**1,633**	**1,595**	**1,595**	**1,564**	**1,616**	1,670	1,799	**1,612**
				Y	24,059	24,721	10,482	**9,982**	15,135	12,387	24,033	**10,346**
200	Yes	100	100	N	2,122	**2,074**	**2,074**	**2,004**	2,107	2,121	**2,016**	2,198
				Y	35,044	36,524	13,969	**12,836**	25,790	13,897	35,069	14,918
100	No	200	100	N	**1,571**	**1,576**	**1,576**	**1,543**	**1,573**	**1,595**	1,697	**1,560**
				Y	23,837	23,925	10,559	**10,040**	15,644	11,616	23,525	**10,219**
100	Yes	200	100	N	2,601	2,215	2,215	**1,863**	2,035	2,126	2,959	2,567
				Y	43,450	50,832	16,362	**14,038**	33,233	**13,426**	43,356	16,467
100	No	100	100	N	**1,698**	**1,756**	**1,756**	**1,705**	**1,767**	**1,767**	**1,698**	**1,712**
				Y	26,476	27,425	**15,028**	**14,771**	19,705	16,167	26,476	**14,859**
100	Yes	100	100	N	2,473	2,336	2,336	**2,103**	2,256	2,318	2,473	2,509
				Y	41,784	44,940	22,037	20,386	33,758	**18,856**	41,784	20,598
200	No	100	100	N	**1,572**	**1,564**	**1,564**	**1,532**	**1,580**	1,636	1,708	**1,561**
				Y	22,840	23,749	**12,954**	**12,911**	16,410	14,599	22,991	**13,050**
200	Yes	100	100	N	2,021	**2,001**	**2,001**	**1,916**	2,026	2,045	**1,944**	2,100
				Y	33,070	35,620	**17,110**	**17,053**	25,926	**16,524**	33,498	17,948
100	No	200	100	N	**1,568**	**1,564**	**1,564**	**1,527**	**1,567**	**1,563**	1,709	**1,553**
				Y	22,943	23,425	**13,469**	**13,304**	16,744	14,460	22,765	**13,241**
100	Yes	200	100	N	2,390	1,985	1,985	**1,784**	1,963	2,006	2,710	2,347
				Y	41,637	46,585	20,809	20,101	32,577	**18,228**	41,359	19,790

NOTE: Table entries are the square root of the EMSE for each estimator of the population size N (first row) and the population total Y (second row). For runs with Clus = Yes, the sample from frame A_1 was drawn as a cluster sample. The entries above the double rule used samples with no misclassification; the entries below the double rule had approximately 10% of the observations classified in the wrong domain. The values within 5% of the minimum $\sqrt{\text{EMSE}}$ in each row are displayed in boldface.

Table 2. Simulation Results for the Population With $\mathbf{N}^T$ = (30,000, 30,000, 10,000, 10,000, 30,000, 10,000, 30,000) and $\boldsymbol{\mu}^T$ = (5, 5, 5, 5, 0, 0, 0)

$\tilde{n}^{(1)}$	Clus?	$n^{(2)}$	$n^{(3)}$	Est.	Ave	Hart	FB	PML	PML12	PML23	SF	SFrake
100	No	100	100	N	**3,417**	3,602	3,602	**3,415**	**3,486**	**3,448**	**3,441**	**3,430**
				Y	20,279	14,553	**13,372**	**12,938**	15,165	**12,985**	20,836	**12,913**
100	Yes	100	100	N	5,290	5,880	5,880	**4,235**	4,515	**4,324**	5,466	5,463
				Y	31,433	23,187	18,712	**16,977**	21,095	**16,891**	32,384	**17,033**
200	No	100	100	N	**3,181**	**3,189**	**3,189**	**3,056**	**3,127**	**3,149**	3,440	**3,126**
				Y	17,841	**12,388**	**12,136**	**11,837**	13,406	**11,827**	15,231	**11,813**
200	Yes	100	100	N	4,376	4,408	4,408	**4,030**	**4,153**	**4,073**	**4,114**	4,585
				Y	24,854	17,527	15,451	**14,547**	17,769	**14,526**	21,073	**15,095**
100	No	200	100	N	**3,058**	**3,080**	**3,080**	**2,969**	**3,029**	**3,010**	3,304	**3,023**
				Y	18,241	14,197	**12,417**	**12,081**	13,511	**12,078**	20,599	**12,050**
100	Yes	200	100	N	5,233	5,488	5,488	**3,481**	3,952	**3,513**	6,224	5,147
				Y	30,507	23,478	18,108	**16,276**	19,588	**16,222**	35,543	**16,288**
100	No	100	100	N	4,113	**3,888**	**3,888**	4,084	4,181	4,087	4,179	4,187
				Y	20,132	17,692	15,895	**14,498**	16,338	**14,609**	20,606	**14,443**
100	Yes	100	100	N	5,607	4,958	4,958	**4,712**	**4,932**	**4,890**	5,756	5,777
				Y	30,885	24,766	21,215	**19,340**	22,986	**19,252**	31,681	**19,162**
200	No	100	100	N	3,810	**3,584**	**3,584**	3,805	3,825	3,792	3,993	3,879
				Y	18,109	16,696	15,901	**14,553**	15,312	**14,643**	16,269	**14,536**
200	Yes	100	100	N	4,706	**4,327**	**4,327**	**4,452**	4,581	4,546	4,580	4,927
				Y	23,153	19,589	17,481	**16,438**	18,652	**16,456**	20,438	**16,998**
100	No	200	100	N	3,805	**3,563**	**3,563**	3,768	3,806	3,775	3,995	3,858
				Y	18,416	17,462	14,758	**13,625**	14,531	**13,631**	20,445	**13,569**
100	Yes	200	100	N	5,526	4,327	4,327	**4,111**	**4,248**	4,352	6,358	5,646
				Y	29,428	23,932	18,322	**17,244**	19,783	**17,209**	33,832	**16,861**

NOTE: The table entries are the square root of the EMSE for each estimator of the population size N (first row) and the population total Y (second row). For runs with Clus = Yes, the sample from frame 1 was drawn as a cluster sample. The observations below the double rule were obtained using misclassified data. The values within 5% of the minimum $\sqrt{\text{EMSE}}$ in each row are displayed in boldface.

estimator with small sample sizes led it to sometimes have higher EMSE than the estimator that simply averages estimates over the three frames. This can be seen in, for example, the fourth line in Table 1, where $\sqrt{\text{EMSE}}$ of the Hartley method for estimating Y is 50,779, whereas $\sqrt{\text{EMSE}}$ of the averaging method is 44,767. If the simulation is rerun with a sample size of 1,000 in each frame, the Hartley method is more stable and has smaller EMSE than the averaging method; its EMSE is still larger than that from the PML method, however.

The full PML method also performed better than the collapsed domain modification described in Section 4.3. When a cluster sample was taken from frame A_1, the EMSE from the PML12 method could be more than twice as large as the EMSE from the full PML method. The EMSE's from the PML23

Table 3. Simulation Results for the Population With $\mathbf{N}^T$ = (30,000, 30,000, 10,000, 10,000, 30,000, 10,000, 10,000) and $\boldsymbol{\mu}^T$ = (−10, 0, 0, 0, 0, 0, 0)

$\tilde{n}^{(1)}$	Clus?	$n^{(2)}$	$n^{(3)}$	Est.	Ave	Hart	FB	PML	PML12	PML23	SF	SFrake
100	No	100	100	N	**3,142**	**3,152**	**3,152**	**3,035**	**3,177**	**3,102**	3,386	**3,049**
				Y	41,427	42,905	32,418	**30,283**	33,732	**30,781**	41,398	**30,904**
100	Yes	100	100	N	5,098	4,738	4,738	**3,655**	3,908	**3,777**	5,666	4,863
				Y	86,862	104,637	56,114	**39,782**	53,939	**40,842**	86,794	57,654
200	No	100	100	N	2,843	**2,730**	**2,730**	**2,662**	**2,764**	**2,736**	3,186	**2,747**
				Y	29,197	29,437	**26,359**	**25,420**	27,276	**25,772**	29,218	**25,576**
200	Yes	100	100	N	4,199	3,816	3,816	**3,576**	3,793	**3,643**	4,125	4,134
				Y	64,456	70,465	40,892	**36,425**	43,242	**37,060**	64,569	45,803
100	No	200	100	N	2,694	**2,616**	**2,616**	**2,557**	**2,635**	**2,574**	3,010	**2,605**
				Y	39,588	40,856	**27,198**	**25,989**	29,062	**26,151**	39,461	27,696
100	Yes	200	100	N	5,027	4,511	4,511	**2,974**	3,230	**3,004**	6,233	4,548
				Y	88,411	109,703	51,298	**32,540**	49,930	**32,870**	88,235	54,422
100	No	100	100	N	3,400	**3,099**	**3,099**	**3,069**	**3,123**	**3,161**	3,878	**3,196**
				Y	42,214	43,904	**35,120**	**33,490**	35,698	**33,704**	43,085	**34,630**
100	Yes	100	100	N	5,048	4,140	4,140	**3,754**	3,951	**3,876**	5,676	4,795
				Y	85,527	101,899	57,911	**47,345**	58,085	**48,260**	85,359	61,823
200	No	100	100	N	3,063	**2,649**	**2,649**	**2,661**	**2,668**	2,833	3,485	2,824
				Y	32,632	33,594	27,696	**25,923**	27,463	**26,415**	31,554	**26,647**
200	Yes	100	100	N	3,970	**3,471**	**3,471**	**3,336**	**3,478**	**3,464**	4,068	3,851
				Y	62,434	76,572	46,830	**40,300**	46,309	**40,967**	62,453	47,690
100	No	200	100	N	2,998	**2,663**	**2,663**	**2,663**	**2,678**	**2,713**	3,394	2,803
				Y	41,281	43,318	**31,963**	**30,498**	32,842	**30,776**	43,062	32,741
100	Yes	200	100	N	4,749	3,357	3,357	**2,876**	**2,969**	**2,984**	5,818	4,345
				Y	84,567	99,537	49,564	**39,629**	53,980	**40,477**	84,130	58,005

NOTE: Table entries are the square root of the EMSE for each estimator of the population size N (first row) and the population total Y (second row). For runs with Clus = Yes, the sample from frame 1 was drawn as a cluster sample. The observations below the double rule were obtained using misclassified data. The values within 5% of the minimum $\sqrt{\text{EMSE}}$ in each row are displayed in boldface.

Table 4. Simulation Results for Population With $\mathbf{N}^T = (10{,}000, 10{,}000, 10{,}000, 10{,}000, 10{,}000, 10{,}000, 10{,}000)$ and $\boldsymbol{\mu}^T = (0, 0, 0, 0, 0, 0, 0)$

$\tilde{n}^{(1)}$	Clus?	$n^{(2)}$	$n^{(3)}$	Est.	Ave	Hart	FB	PML	PML12	PML23	SF	SFrake
100	No	100	100	Y	**6,197**	**6,249**	**6,431**	**6,230**	**6,233**	**6,269**	**6,197**	**6,207**
100	Yes	100	100	Y	**8,050**	**8,060**	8,860	**7,931**	**7,868**	**7,734**	**8,050**	**8,069**
200	No	100	100	Y	**5,613**	**5,628**	**5,747**	**5,567**	**5,575**	**5,604**	**5,566**	**5,552**
200	Yes	100	100	Y	**6,730**	**6,818**	7,137	**6,869**	**6,854**	**6,769**	**6,930**	**6,946**
100	No	200	100	Y	**5,552**	**5,479**	**5,617**	**5,468**	**5,522**	**5,500**	**5,440**	**5,440**
100	Yes	200	100	Y	7,687	7,775	8,233	**7,471**	**7,465**	**7,308**	**7,461**	**7,405**
100	No	100	100	Y	**6,119**	**6,182**	**6,368**	**6,148**	**6,139**	**6,178**	**6,119**	**6,122**
100	Yes	100	100	Y	**7,829**	**7,732**	8,282	**7,537**	**7,595**	**7,495**	**7,829**	**7,827**
200	No	100	100	Y	**5,642**	**5,607**	**5,723**	**5,599**	**5,611**	**5,606**	**5,576**	**5,575**
200	Yes	100	100	Y	**6,664**	**6,706**	6,996	**6,653**	**6,683**	**6,650**	**6,931**	**6,946**
100	No	200	100	Y	**5,773**	**5,727**	**5,856**	**5,729**	**5,787**	**5,731**	**5,693**	**5,705**
100	Yes	200	100	Y	7,472	**7,236**	7,668	**7,031**	**7,111**	**6,951**	**7,153**	**7,111**

NOTE: Table entries are the square root of the EMSE for each estimator of the population total Y. For runs with Clus = Yes, the sample from frame 1 was drawn as a cluster sample. The observations below the double rule were obtained using misclassified data. The values within 5% of the minimum $\sqrt{\text{EMSE}}$ in each row are displayed in boldface. Simulation results for estimating N for these settings were given in Table 1.

method were in general smaller than those from PML12 when a cluster sample was taken from frame A_1, because the simple random samples in frames A_2 and A_3 had larger effective sample sizes for estimating the population size of $A_2 \cup A_3$. The PML12 and PML23 estimates were also much more complicated and time-consuming to compute than the full PML method, due to the extra steps involved in estimating variances of collapsed domains. When samples of size 1,000 or more were taken in each frame, the PML12 and PML23 methods performed similarly to the full PML, but in this case there is no reason not to use the computationally simpler full PML method.

The single-frame estimator with no raking adjustment performed poorly in many of the simulations. In some cases, increasing the sample size in one of the frames increased, rather than decreased, the EMSE of the single-frame estimator. This phenomenon can be seen in Tables 1–3 when a cluster sample with $\tilde{n}^{(1)} = 100$ was taken in frame A_1; increasing $n^{(2)}$ from 100 to 200 increased $V(\hat{N}_S)$. The single-frame estimator for the sampling designs used in this simulation has the form $\sum_{q=1}^{Q} \boldsymbol{\theta}_q^T \boldsymbol{\Gamma}_q \hat{\mathbf{Y}}_q$ as in (4), but the values of $\boldsymbol{\theta}_q$ are functions solely of the frame weights and do not adjust for the relative variances of the domain estimates. This same effect can be demonstrated analytically for a dual-frame design with both frames incomplete; if $N_{\{1\}} = N_{\{2\}} = N_{\{1,2\}}$ and $N^{(1)} = N^{(2)}$, then it can be shown that, ignoring finite-population corrections,

$$V(\hat{N}_S) = \frac{N_{\{1\}} N_{\{1,2\}}}{[N^{(1)}(f^{(1)} + f^{(2)})]^2} \left[\frac{m(f^{(2)})^2}{f^{(1)}} + \frac{(f^{(1)})^2}{f^{(2)}} \right],$$

where m is the cluster size in the sample from frame A_1. Thus increasing $f^{(2)}$ while holding $f^{(1)}$ constant can indeed increase the variance of the single-frame estimator.

In most of the simulation runs, the raking ratio single-frame estimator had much smaller EMSEs than the unadjusted single-frame estimator. However, the EMSE from the single-frame raking estimator was generally higher than that of the PMLE when a cluster sample was taken from frame 1. The raking ratio estimator exhibited its best performance in Tables 2 and 4; at best, it performed comparably to the PML method.

When 9.75% of the observations were classified in the wrong domain, the same general patterns for the magnitudes of the EMSEs held as in the correct classification case. All estimators are approximately unbiased when observations are classified in the correct domain, but misclassification can result in biased estimators. We had conjectured that the collapsed-domain PML methods from Section 4.3 might exhibit less bias than the full PML method, but this turned out not to be the case. No method consistently exhibited more or less bias than the others. For this simulation study, most of the changes in EMSE from misclassification were due to changes in the variance, not to the bias. The variances for the methods appeared to be differentially affected by this misclassification scheme; the variance of the PML methods often increased by a higher percentage than the variance of the H or FB methods. The effects of misclassification also differed depending on whether a simple random sample or a cluster sample was drawn from frame A_1. With a simple random sample from frame A_1, the EMSEs with misclassification either increased or remained about the same. With a cluster sample, however, some of the EMSE's decreased with misclassification; in this simulation study individual observations, not clusters, were misclassified, so that misclassification led to a smaller design effect in frame A_1. More research on effects of misclassification on the variances of estimates is needed; in this study, however, the PML method exhibited the smallest EMSEs even with misclassification.

In all of the simulations, the PML method had either the smallest EMSE or an EMSE close to the minimum value. With its high efficiency and ease of computation, as well as the practical advantage of using the same set of weights for all response variables, the PML method appears to be a good choice for estimation in multiple-frame surveys.

8. VARIANCE ESTIMATION

Several methods have been developed for finding variance estimators for dual-frame estimators. Lohr and Rao (2000) described Taylor linearization and jackknife methods for estimating variances of functions of estimated population totals in dual-frame surveys. Those methods can be extended to surveys with more than two frames. To avoid confusion with previous notation, in this section let $\mathbf{X}^{(q)}$ denote an $r^{(q)}$-vector of population totals in frame A_q and let $\bar{\mathbf{X}}^{(q)} = \mathbf{X}^{(q)}/N^{(q)}$ denote the corresponding vector of population means. Let $\hat{\mathbf{X}}^{(q)}$ be an unbiased estimator of $\mathbf{X}^{(q)}$.

We are interested in estimating a parameter

$$\tau = g(\bar{\mathbf{X}}^{(1)}, \dots, \bar{\mathbf{X}}^{(Q)})$$

and finding the standard error of the estimator

$$\hat{\tau} = g(\hat{\bar{\mathbf{X}}}^{(1)}, \dots, \hat{\bar{\mathbf{X}}}^{(Q)}). \quad (24)$$

All of the estimators of population totals in Sections 3–5 can be expressed in the form of (24). The generalized Hartley estimator $\hat{Y}_{\boldsymbol{\theta}}$ in (7) may be expressed as $\hat{\mathbf{Y}}_{\boldsymbol{\theta}} = g(\hat{\bar{\mathbf{X}}}^{(1)}, \dots, \hat{\bar{\mathbf{X}}}^{(Q)})$, with $g(\mathbf{x}_1, \dots, \mathbf{x}_Q) = \sum_{q=1}^{Q} N^{(q)} \boldsymbol{\theta}_q^T \boldsymbol{\Gamma}_q \mathbf{x}_q$, and $\hat{\bar{\mathbf{X}}}^{(q)} = \hat{Y}_{(q)}/N^{(q)}$. The function g is defined similarly for $\hat{Y}_{FB}$ in (9), with $\hat{\bar{\mathbf{X}}}^{(q)} = (\hat{Y}_{(q)}^T/N^{(q)}, \hat{N}_{(q)}^T/N^{(q)})$. For the single-frame estimator in (23), $\hat{Y}_S = g(\hat{\bar{\mathbf{X}}}^{(1)}, \dots, \hat{\bar{\mathbf{X}}}^{(Q)})$ with $g(\mathbf{x}_1, \dots, \mathbf{x}_Q) = \sum_{q=1}^{Q} N^{(q)} \mathbf{x}_q$ and $\hat{\bar{\mathbf{X}}}^{(q)} = \sum_{i \in \mathcal{S}_q} \tilde{w}_{i,S}^{(q)} y_i / N^{(q)}$. The function g is implicitly defined for the PMLE.

Suppose that for $q = 1, \dots, Q$, frame A_q has $H^{(q)}$ strata and stratum $h^{(q)}$ has $N_h^{(q)}$ observation units and $\tilde{N}_h^{(q)}$ primary sampling units (psu's), of which $\tilde{n}_h^{(q)}$ are sampled. Stratified simple random sampling in one or more of the frames, say a list frame A_2, follows as a special case with individual observations in a stratum comprising the psu's. Define $\tilde{n}^{(q)} = \sum_{h=1}^{H^{(q)}} \tilde{n}_h^{(q)}$ and $W_h^{(q)} = N_h^{(q)}/N^{(q)}$. Let $\mathbf{S}^{(q)}$ be an estimate of $V(\hat{\bar{\mathbf{X}}}^{(q)})$ and let $g_q(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_Q)$ be the $r^{(q)}$-vector of first derivatives of g with respect to the components of $\mathbf{x}_q$, evaluated at $(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_Q)$. The linearization variance estimator is then defined as

$$v_L(\hat{\tau}) = \sum_{q=1}^{Q} g_q^T(\hat{\bar{\mathbf{X}}}^{(1)}, \dots, \hat{\bar{\mathbf{X}}}^{(Q)}) \mathbf{S}^{(q)} g_q(\hat{\bar{\mathbf{X}}}^{(1)}, \dots, \hat{\bar{\mathbf{X}}}^{(Q)}). \quad (25)$$

To define the jackknife estimator of the variance, let $\hat{\tau}_{(hi)}^{(q)}$ be the estimator of the same form as $\hat{\tau}$ when the observations of sample psu i of stratum h are omitted from the sample from frame q, $\hat{\tau}_{(hi)}^{(q)} = g(\hat{\bar{\mathbf{X}}}^{(1)}, \dots, \hat{\bar{\mathbf{X}}}^{(q-1)}, \hat{\bar{\mathbf{X}}}_{(hi)}^{(q)}, \hat{\bar{\mathbf{X}}}^{(q+1)}, \dots, \hat{\bar{\mathbf{X}}}^{(Q)})$, where $\hat{\bar{\mathbf{X}}}_{(hi)}^{(q)}$ is the estimator of $\bar{\mathbf{X}}^{(q)}$ computed after omitting sample psu i of stratum h in frame q. A jackknife variance estimator of $\hat{\tau}$ is then given by

$$v_J(\hat{\tau}) = \sum_{q=1}^{Q} \sum_{h=1}^{H^{(q)}} \frac{\tilde{n}_h^{(q)} - 1}{\tilde{n}_h^{(q)}} \sum_{i=1}^{\tilde{n}_h^{(q)}} (\hat{\tau}_{(hi)}^{(q)} - \hat{\tau})^2. \quad (26)$$

Assuming with-replacement sampling with

$$\tilde{n}^{(q)} \Big/ \left(\sum_{q=1}^{Q} \tilde{n}^{(q)} \right) \to k_q \in (0, 1)$$

for $q = 1, \dots, Q$, we have $v_L(\hat{\tau})/V(\hat{\tau}) \to_P 1$ and $v_J(\hat{\tau})/v_L(\hat{\tau}) \to_P 1$ under regularity conditions similar to those in theorem 3 of Lohr and Rao (2000). Various extensions of the jackknife variance estimator in (26) to cover more complex situations may be obtained along the lines of Lohr and Rao (2000) for dual-frame surveys. Bootstrap methods may also be used to estimate variances, and this is the subject of current research.

9. APPLICATION

As an illustration of the method, data were collected for a three-frame survey of statisticians, using the online membership directories of the American Statistical Association (frame A_1), the Institute for Mathematical Statistics (frame A_2), and the Statistical Society of Canada (frame A_3). At the time of data collection, $N_1 = 15{,}500$, $N_2 = 4{,}000$, and $N_3 = 863$. Admittedly, the union of the three frames does not cover the entire population of statisticians; many statisticians do not belong to any of the three societies, and other statisticians decline to participate in online directories.

A stratified cluster sample of size $n_1 = 500$ was taken from frame A_1; the 26 strata used were regions or states, and, because of the restrictions on access to records, clusters for large states were members whose last name began with the same letter of the alphabet. Disproportional allocation was used, with weights ranging from 16 (in the District of Columbia) to about 40. For frames A_2 and A_3, independent simple random samples of sizes 140 and 150 were selected.

Each of frames A_1, A_2, and A_3 is incomplete, and all three frames overlap, indicating the situation in Figure 1(a). Closed-form solutions for the estimators in this design are complicated, and we used the iterative approach described in Section 4.1. The matrix $\mathbf{M}$ and its Moore–Penrose generalized inverse $\mathbf{M}^+$, corresponding to the ordering

$$\mathbf{N}^T = (N_{\{1\}}, N_{\{1,2\}}, N_{\{1,3\}}, N_{\{1,2,3\}}, N_{\{2\}}, N_{\{2,3\}}, N_{\{3\}}),$$

are given by

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and}$$

$$\mathbf{M}^+ = \frac{1}{8} \begin{bmatrix} 3 & 2 & 2 & 1 & -1 & -2 & -1 \\ -1 & 2 & -2 & 1 & 3 & 2 & -1 \\ -1 & -2 & 2 & 1 & -1 & 2 & 3 \end{bmatrix}.$$

From the samples,

$$\hat{\mathbf{H}} = \begin{bmatrix} 12{,}389 & 0 & 0 \\ 1{,}721 & 1{,}600 & 0 \\ 112 & 0 & 173 \\ 168 & 57 & 167 \\ 0 & 2{,}257 & 0 \\ 0 & 86 & 58 \\ 0 & 0 & 466 \end{bmatrix}.$$

A substantial fraction of the population belongs to exactly one of the frames, so the multiple-frame survey provides better coverage of the population than a single-frame survey.

Equation (18) gives a polynomial equation of degree 6. The linearized estimator of $\mathbf{N}$ in (19) is considerably simpler and can be easily solved for any specific values of $\mathbf{f}$ and the $\hat{N}_K^{(q)}$'s. Using the averages of the domain estimates over the frames as initial estimates gives $\tilde{\mathbf{N}}_0 = [12{,}389, 1{,}660.5, 142.5, 130.7, 2{,}257, 72, 466]^T$. Simple random samples were taken in frames

A_2 and A_3, so that $f^{(2)} = .035$ and $f^{(3)} = .1738$. To determine $f^{(1)}$, we used the median of the design effects for estimating $N_{\{1\}}, N_{\{1,2\}}, N_{\{1,3\}}$, and $N_{\{1,2,3\}}$, which was 2.1, and set $f^{(1)} = (1/2.1)(500/15{,}500) = .015$. The iterations with (20) converged with $\tilde{\mathbf{N}}_5 = \hat{\mathbf{N}}_{\text{PML}} = [13{,}548, 1{,}628, 172, 152, 2{,}157, 63, 476]^T$.

We estimated the size of the union of the three frames (N) and the proportion of women in the population (p), using the methods described in Sections 3 and 4. We did not use the single-frame method described in Section 5, because the frame A_1 weight is unknown for sampled units in the other two frames. The estimates are given in the following table; standard errors, in parentheses below the estimates, were calculated using a modification of the jackknife in which psu's were randomly split into two pseudo-psu's when only one psu appeared in a stratum:

Estimate	*Average*	*Hartley*	*FB*	*PML*	*PML12*	*PML23*
$\hat{N}$	17,118	16,875	16,875	18,196	18,217	18,190
	(1,011)	(895)	(895)	(144)	(152)	(165)
$\hat{p}$	.280	.286	.288	.284	.283	.284
	(.023)	(.020)	(.020)	(.021)	(.021)	(.021)

All methods gave similar estimates for p. For estimating the population size N, the PML methods had much smaller standard errors, in part because they used the design effects along with auxiliary information on the frame population sizes.

10. DISCUSSION

As the United States, Canada, and other nations grow in diversity, different sampling frames may better capture subgroups of the population. Some population subgroups are not well covered by traditional frames, and a multiple-frame approach may help reduce undercoverage biases as well as costs of surveys of a diverse population. One advantage of multiple-frame methods is that large national surveys may be easily supplemented with additional data from a state or province or for a subgroup of the population. We anticipate that modular sampling designs using multiple frames will be widely used in the future.

In this article we have presented methods that may be used to combine information from multiple frames that are subsets of the same population. Stukel, Mohl, and Tambay (1997) discussed the advantages of the Skinner and Rao (1996) PMLEs in the Canadian National Population Health Survey. The PMLE presented in this article shares these advantages; it uses the same weights for all variables and is nearly optimal for estimating Y if the design effects for $\hat{Y}^{(q)}$ are close to those used in setting $f^{(q)}$. The single-frame estimator with raking ratio adjustments also performed well for most cases in the simulation study, but with cluster sampling sometimes had an MSE substantially higher than that of the PMLE.

The other estimators studied in this article did not perform as well as either the PMLE or the single-frame estimator with raking ratio adjustment, and we do not recommend their use in practice with more than two frames. When $Q \geq 3$, the theoretically optimal Fuller–Burmeister and Hartley methods became unstable, because they require solving systems of equations using a large estimated covariance matrix. The unraked single-frame estimator performed poorly relative to the PMLE in almost every simulation run; it also has the paradoxical property that when cluster samples are taken in one or more frames, increasing the sample size in one of the frames can lead to an increased variance of the estimators. This same phenomenon can occur with the raked single-frame estimator as well, but it is not as common and generally does not result in a large increase in the variance. The combined PMLEs described in Section 4.3 performed better than the unraked single-frame estimator and comparably to the raked single-frame estimator, but not as well as the full PML method; they are also much more complicated to compute than the raked single-frame and PMLE methods. Among the estimators studied, we recommend the multiframe PMLE as a good choice for a wide variety of conditions.

Multiple-frame surveys can greatly improve efficiency and reduce undercoverage bias, but must be used carefully. Multiple-frame estimation methods implicitly assume that differences among estimates from the various frames are due to sampling error and the subset of the population in the frame. If samples taken from the different frames use different questionnaires or modes of administration, then care must be taken that these do not create biases. All of the methods used for estimation with overlapping frames presented in this article also assume that domain membership can be determined for every sampled unit. These methods are sensitive to misclassification of observations into domains, and estimates can be biased with misclassification if the domain means differ. Sampling designs for multiple-frame surveys should consider nonresponse errors and misclassification, as well as variances of estimators.

APPENDIX: PROOFS

Proof of (18)

Let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_Q)^T$. Maximizing (15) subject to the constraint in (16) is equivalent to maximizing

$$g(\mathbf{N}, \boldsymbol{\lambda}) = \sum_{q=1}^{Q} \sum_{K:\, q \in K} \hat{N}_K^{(q)} f^{(q)} \log(N_K) + \sum_{q=1}^{Q} \lambda_q \left(N^{(q)} - \sum_{K:\, q \in K} N_K \right).$$

Now

$$\frac{\partial g}{\partial N_K} = \sum_{q \in K} \frac{\hat{N}_K^{(q)} f^{(q)}}{N_K} - \boldsymbol{\lambda}^T \mathbf{1}_Q$$

and

$$\frac{\partial g}{\partial \lambda_q} = N^{(q)} - \sum_{K:\, q \in K} N_K.$$

In matrix notation, the score equations are

$$\begin{bmatrix} (\operatorname{diag} \hat{\mathbf{N}})^{-1} \hat{\mathbf{H}} \mathbf{f} - \mathbf{M} \boldsymbol{\lambda} \\ \mathbf{M}^T \hat{\mathbf{N}} - \mathbf{h} \end{bmatrix} = \mathbf{0}_{Q+d}.$$

Equation (18) results by noting that $\mathbf{M}\boldsymbol{\lambda} = \mathbf{M}\mathbf{M}^{+}\mathbf{M}\boldsymbol{\lambda}$.

Proof of (19)

Note that $\hat{\mathbf{H}}_{ij}$ estimates $\mathbf{H}$, where $\mathbf{H}_{ij} = N_{K_i}$ if $j \in K_i$ and 0 otherwise. Let $\mathbf{L}$ be a $(d - Q) \times d$ matrix of linearly independent rows of $(\mathbf{I} - \mathbf{M}\mathbf{M}^{+})$. Then $\hat{\mathbf{N}}$ is defined implicitly as $\boldsymbol{\eta}(\hat{\mathbf{H}}, \hat{\mathbf{N}}) = \mathbf{0}_d$, where

$$\boldsymbol{\eta}(\mathbf{A}, \mathbf{z}) = \begin{bmatrix} \mathbf{L} (\operatorname{diag} \mathbf{z})^{-1} \mathbf{A} \mathbf{f} \\ \mathbf{M}^T \mathbf{z} - \mathbf{h} \end{bmatrix}.$$

Note that $\boldsymbol{\eta}(\mathbf{H}, \mathbf{N}) = \mathbf{0}_d$. Using the implicit mapping theorem, the first-step approximation to $\hat{\mathbf{N}}$ is

$$\tilde{\mathbf{N}} = \mathbf{N} - \left[\frac{\partial \boldsymbol{\eta}(\mathbf{A}, \mathbf{z})}{\partial \mathbf{z}^T} \bigg|_{\mathbf{A}=\mathbf{H}, \mathbf{z}=\mathbf{N}} \right]^{-1} \boldsymbol{\eta}(\hat{\mathbf{H}}, \mathbf{N}).$$

The proof of (19) is completed by showing that

$$\left[\frac{\partial \boldsymbol{\eta}(\mathbf{A}, \mathbf{z})}{\partial \mathbf{z}^T} \bigg|_{\mathbf{A}=\mathbf{H}, \mathbf{z}=\mathbf{N}} \right] = \begin{bmatrix} -\mathbf{L}(\operatorname{diag} \mathbf{N})^{-1} \operatorname{diag}[\mathbf{Mf}] \\ \mathbf{M}^T \end{bmatrix}$$

and that

$$\begin{bmatrix} -\mathbf{L}(\operatorname{diag} \mathbf{N})^{-1} \operatorname{diag}[\mathbf{Mf}] \\ \mathbf{M}^T \end{bmatrix} \mathbf{N} = \begin{bmatrix} \mathbf{0}_{d-Q} \\ \mathbf{h} \end{bmatrix}.$$

[Received June 2004. Revised December 2005.]

REFERENCES

Bankier, M. D. (1986), "Estimators Based on Several Stratified Samples With Applications to Multiple Frame Surveys," *Journal of the American Statistical Association*, 81, 1074–1079.

Chu, A., Brick, J. M., and Kalton, G. (1999), "Weights for Combining Surveys Across Time or Space," paper presented at the 52nd Meeting of the International Statistical Institute, Helsinki.

Fuller, W. A., and Burmeister, L. F. (1972), "Estimators for Samples Selected From Two Overlapping Frames," in *Proceedings of the Social Statistics Section*, American Statistical Association, pp. 245–249.

González-Villalobos, A., and Wallace, M. A. (1996), *Multiple-Frame Agriculture Surveys*, Vols. 1 and 2, Rome: Food and Agriculture Organization of the United Nations.

Haines, D. E., and Pollock, K. H. (1998), "Combining Multiple Frames to Estimate Population Size and Totals," *Survey Methodology*, 24, 79–88.

Hartley, H. O. (1962), "Multiple Frame Surveys," in *Proceedings of the Social Statistics Section*, American Statistical Association, pp. 203–206.

——— (1974), "Multiple Frame Methodology and Selected Applications," *Sankhyā*, Ser. C, 36, 99–118.

Iachan, R., and Dennis, M. L. (1993), "A Multiple Frame Approach to Sampling the Homeless and Transient Population," *Journal of Official Statistics*, 9, 747–764.

Kalton, G., and Anderson, D. W. (1986), "Sampling Rare Populations," *Journal of the Royal Statistical Society*, Ser. A, 149, 65–82.

Lohr, S. L., and Rao, J. N. K. (2000), "Inference in Dual Frame Surveys," *Journal of the American Statistical Association*, 95, 271–280.

Madans, J. H., Ezzati-Rice, T. M., Cynamon, M., and Blumberg, S. J. (2001), "Targeting Approaches to State-Level Estimates," in *Proceedings of the Seventh Conference on Health Survey Research Methods*, eds. M. L. Cynamon and R. A. Kulka, Hyattsville, MD: Department of Health and Human Services, pp. 239–245.

National Agricultural Statistics Service (1999), "U.S. Equine Inventory Up 1.3 Percent," news release, March 2, 1999; available at *http://www.usda.gov/nass/pubs/pubs.htm*.

O'Muircheartaigh, C., and Pedlow, S. (2002), "Combining Samples versus Cumulating Cases: A Comparison of Two Weighting Strategies in NLSY97," in *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 2557–2562.

Rao, J. N. K. (2003), *Small-Area Estimation*, New York: Wiley.

Rao, J. N. K., and Skinner, C. J. (1996), "Estimation in Dual Frame Surveys With Complex Designs," in *Proceedings of the Survey Methods Section*, Statistical Society of Canada, pp. 63–68.

Schoorl, J., Heering, L., Esveldt, I., Groenewold, G., van der Erf, R., Bosch, A., de Valk, H., and de Bruijn, B. (2000), *Push and Pull Factors of International Migration: A Comparative Report*, Luxembourg: Office for Official Publications of the European Communities.

Skinner, C. J. (1991), "On the Efficiency of Raking Ratio Estimation for Multiple Frame Surveys," *Journal of the American Statistical Association*, 86, 779–784.

Skinner, C. J., Holmes, D. J., and Holt, D. (1994), "Multiple Frame Sampling for Multivariate Stratification," *International Statistical Review*, 62, 333–347.

Skinner, C. J., and Rao, J. N. K. (1996), "Estimation in Dual Frame Surveys With Complex Designs," *Journal of the American Statistical Association*, 91, 349–356.

Stukel, D. M., Mohl, C. A., and Tambay, J.-L. (1997), "Weighting for Cycle Two of Statistics Canada's National Population Health Survey," in *Proceedings of the Survey Methods Section*, Statistical Society of Canada, pp. 111–116.