JACKKNIFE VARIANCE ESTIMATION IN MULTIPLE FRAME SURVEYS

Sharon Lohr, Arizona State University and J.N.K. Rao, Carleton University
Sharon Lohr, Department of Mathematics, Arizona State Univ., Tempe AZ 85287-1804

Key Words: Complex surveys, Dual frame surveys, Linearization variance estimation, Resampling methods.

Abstract:

In multiple frame surveys, samples are drawn independently from overlapping frames that together cover the population of interest. We propose jackknife variance estimators for multiple frame surveys and establish their design consistency.

1. Introduction

In a dual frame survey, two sampling frames, A and B, together cover the population of interest $\mathcal{U}$. Independent probability samples are taken from frames A and B, and information from the two samples is combined to estimate quantities of interest.

A common use of dual frame surveys occurs when frame A is an area frame and frame B a list frame. Frame A is complete but expensive to sample; frame B, while incomplete, also has a lower cost per unit sampled. For example, frame A might be an area code/prefix frame used with random digit dialing in a telephone survey, and frame B a commercial directory of residential telephone numbers. Frame A contains all telephone numbers in the population, but has no auxiliary information that can be used to design an efficient sampling scheme; most of the numbers in frame B are from residential households and the frame includes additional information such as address, but not all residential households appear in frame B. Vogel (1975) discusses some applications in agricultural surveys.

Let y_i be a measurement from observation unit i. A number of estimators have been proposed for estimating the population total $Y = \sum_{i \in \mathcal{U}} y_i$ by Hartley (1962, 1974), Fuller and Burmeister (1972), Lund (1968), Bankier (1986), Kalton and Anderson (1986), Skinner (1991), and Skinner and Rao (1996). Most of this work has focussed on the derivation of point estimators $\hat{Y}$ for Y, although Skinner and Rao (1996) mention variance estimators for $\hat{Y}$ based on asymptotic results using linearization. In this paper, theoretical results for a jackknife estimator of the variance are presented, and they are illustrated for the pseudo-maximum likelihood estimator (Skinner and Rao, 1996) of the population total.

This research was partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

2. Estimators of a Total

Following Hartley (1962, 1974), let $\mathcal{A}$ and $\mathcal{B}$ denote the sets of population units in frames A and B, respectively. Then the universe may be divided into three mutually exclusive domains, $\mathcal{U} = a \cup ab \cup b$, where $a = \mathcal{A}\mathcal{B}^c$, $ab = \mathcal{A}\mathcal{B}$, and $b = \mathcal{A}^c\mathcal{B}$. The quantities N, N_A, N_B, N_a, N_b, and N_{ab} are the number of population elements in $\mathcal{U}, \mathcal{A}, \mathcal{B}, a, b$, and ab, respectively. The sample from frame A is denoted $\mathcal{S}_A$, the probability of inclusion in $\mathcal{S}_A$ is $\pi_i^A = P\{i \in \mathcal{S}_A\}$, and $\mathcal{S}_A$ contains n_A observation units. Corresponding quantities for frame B are $\mathcal{S}_B$, $\pi_i^B = P\{i \in \mathcal{S}_B\}$, and n_B.

Let Y_a, Y_{ab}, and Y_b be the population totals, and $\mu_a = Y_a/N_a$, $\mu_{ab} = Y_{ab}/N_{ab}$, and $\mu_b = Y_b/N_b$ be the population means, in domains a, ab, and b respectively. Estimators of $Y = Y_a + Y_{ab} + Y_b$ proposed in the references cited above are all of the form $\hat{Y} = \hat{Y}_a + \hat{Y}_{ab} + \hat{Y}_b$; they differ in how the information from the two samples is combined to obtain estimators of the components Y_a, Y_{ab}, and Y_b.

We assume that N_A and N_B are known, and that $N_a > 0$ and $N_b > 0$. Let w_i^A and w_i^B be the sampling weights for the designs used in frames A and B, respectively. For N_A and N_B known, $w_i^A = N_A[\pi_i^A \sum_{j \in \mathcal{S}_A}(1/\pi_j^A)]^{-1}$ and $w_i^B = N_B[\pi_i^B \sum_{j \in \mathcal{S}_B}(1/\pi_j^B)]^{-1}$. Let $\delta_A(i) = 1$ if $i \in \mathcal{A}$ and 0 otherwise; and $\delta_B(i) = 1$ if $i \in \mathcal{B}$ and 0 otherwise. Then define the domain estimators

$$\begin{aligned}
\hat{N}_a^A &= \textstyle\sum_{i \in \mathcal{S}_A} w_i^A (1 - \delta_B(i)) \\
\hat{N}_{ab}^A &= \textstyle\sum_{i \in \mathcal{S}_A} w_i^A \delta_B(i) \\
\hat{Y}_a^A &= \textstyle\sum_{i \in \mathcal{S}_A} w_i^A (1 - \delta_B(i)) y_i \\
\hat{Y}_{ab}^A &= \textstyle\sum_{i \in \mathcal{S}_A} w_i^A \delta_B(i) y_i
\end{aligned}$$

and

$$\begin{aligned}\hat{N}_b^B &= \textstyle\sum_{i\in S_B} w_i^B(1-\delta_A(i))\\ \hat{N}_{ab}^B &= \textstyle\sum_{i\in S_B} w_i^B\delta_A(i)\\ \hat{Y}_b^B &= \textstyle\sum_{i\in S_B} w_i^B(1-\delta_A(i))y_i\\ \hat{Y}_{ab}^B &= \textstyle\sum_{i\in S_B} w_i^B\delta_A(i)y_i.\end{aligned}$$

Also define

$$\begin{aligned}\hat{Y}_{ab}(\theta) &= \theta\hat{Y}_{ab}^A+(1-\theta)\hat{Y}_{ab}^B\\ \hat{N}_{ab}(\theta) &= \theta\hat{N}_{ab}^A+(1-\theta)\hat{N}_{ab}^B.\end{aligned}$$

Hartley and Fuller-Burmeister Estimators. Hartley (1962) proposed the estimator

$$\hat{Y}_H(\theta)=\hat{Y}_a^A+\hat{Y}_b^B+\hat{Y}_{ab}(\theta),$$

and Fuller and Burmeister (1972) proposed

$$\hat{Y}_{FB}(\beta_1,\beta_2)=\hat{Y}_a^A+\hat{Y}_b^B+\hat{Y}_{ab}(\beta_1)+\beta_2(\hat{N}_{ab}^A-\hat{N}_{ab}^B).$$

The optimal values of the parameters θ, β_1, and β_2 minimize the variance of $\hat{Y}_H(\theta)$ and $\hat{Y}_{FB}(\beta_1,\beta_2)$, and thus depend on covariances of $\hat{Y}_{ab}^A$ and $\hat{Y}_{ab}^B$. In practice, of course, the true covariances are unknown, and must be estimated from the data. As both $\hat{\beta}_{FB}$ and $\hat{\theta}_H$ rely on estimated covariances, the optimal Hartley and Fuller-Burmeister estimators $\hat{Y}_H(\hat{\theta}_H)$ and $\hat{Y}_{FB}(\hat{\beta}_{FB})$ are not in general linear functions of y; a different set of weights would need to be calculated for each response variable. Besides adding to the amount of calculation, the different sets of weights may lead to disagreements among estimates. For example, if $\hat{Y}_{H1}$, $\hat{Y}_{H2}$, and $\hat{Y}_{H3}$ are Hartley estimates for the total number of asthmatics in age groups 0-16, 17-45, and over 45, then $\hat{Y}_{H1}+\hat{Y}_{H2}+\hat{Y}_{H3}$ will not necessarily equal the Hartley estimate of the total number of asthmatics in the entire population.

Pseudo-maximum likelihood estimator. Skinner and Rao (1996) proposed modifying the maximum likelihood estimator for a simple random sample to obtain a pseudo-maximum-likelihood estimator (PML) for a complex design. The PML estimator, unlike the Hartley and Fuller-Burmeister estimators, is linear in y, and is of the following form:

$$\hat{Y}_{PML}(\theta)=\frac{N_A-\hat{N}_{ab}^{PML}(\theta)}{\hat{N}_a^A}\hat{Y}_a^A +\frac{N_B-\hat{N}_{ab}^{PML}(\theta)}{\hat{N}_b^B}\hat{Y}_b^B+\frac{\hat{N}_{ab}^{PML}(\theta)}{\hat{N}_{ab}(\theta)}\hat{Y}_{ab}(\theta), \quad (1)$$

where $\hat{N}_{ab}^{PML}(\theta)$, a function of $\hat{N}_{ab}^A$, $\hat{N}_{ab}^B$, and θ, is the smaller of the roots of the quadratic equation

$$\left(\frac{\theta}{N_B}+\frac{1-\theta}{N_A}\right)x^2 \quad -\left(1+\frac{\theta}{N_B}\hat{N}_{ab}^A+\frac{1-\theta}{N_A}\hat{N}_{ab}^B\right)x +\hat{N}_{ab}(\theta)=0.$$

Skinner and Rao (1996) suggested choosing $\theta=\theta_P$ to minimize the asymptotic variance of $\hat{N}_{ab}^{PML}(\theta)$, with

$$\theta_P=\frac{N_aN_BV(\hat{N}_{ab}^B)}{N_aN_BV(\hat{N}_{ab}^B)+N_bN_AV(\hat{N}_{ab}^A)}. \quad (2)$$

In practice, N_a, N_b, and the variances in (2) are unknown and must be estimated from the data. The resulting estimator is $\hat{Y}_{PML}(\hat{\theta}_P)$. The PML estimator uses the same set of weights for each response variable, and thus avoids some of the difficulties associated with the Hartley and Fuller-Burmeister estimators. If independent simple random samples are taken from frames A and B, the PML estimator is equivalent to the Fuller-Burmeister estimator. Skinner and Rao (1996) found a sufficient condition for the PML estimator to be optimal.

Single Frame Estimators. Bankier (1986), Kalton and Anderson (1986), and Skinner (1991) proposed estimating the population total by treating all observations as though they had been sampled from a single frame with modified weights for observations in the intersection ab. The modified weights for the single-frame estimators of Kalton and Anderson (1986) and Skinner (1991) do not require identification of units found in both samples. These weights are $w_i=1/\pi_i^A$ for $i\in a$, $w_i=1/\pi_i^B$ for $i\in b$, and $w_i=(\pi_i^A+\pi_i^B)^{-1}$ for observations in both frames, so that

$$\hat{Y}_{SF}=\sum_{i\in S_A} w_iy_i+\sum_{i\in S_B} w_iy_i.$$

As Bankier (1986) noted, the single frame estimator may be extended to multiple frame surveys.

The single frame estimator $\hat{Y}_{SF}$ does not use any auxiliary information about the population totals N_A and N_B. It can be adjusted either through raking ratio estimation (Bankier 1986) or regression estimation, as discussed in Rao and Skinner (1996) and Lohr and Rao (1997).

3. Variance Estimation

Skinner and Rao (1996) described a method for estimating the variance of $\hat{Y}_{PML}$ using Taylor linearization. In this section, we define a jackknife variance estimator for estimators from dual frame surveys, and show that the jackknife variance estimator is asymptotically equivalent to the Taylor linearization estimator. We state the results for dual frame surveys to simplify notation; however, the results of this section are easily extended to multiple frame surveys in which independent samples are selected from the frames. We refer the reader to Lohr and Rao (1997) for proofs and further details.

Suppose frame A has H strata and stratum h has N_h^A observation units and $\tilde{N}_h^A$ primary sampling units (psu's), of which $\tilde{n}_h^A$ are sampled. Frame B has L strata and stratum l has N_l^B observation units and $\tilde{N}_l^B$ psu's, of which $\tilde{n}_l^B$ are sampled. Define $\tilde{n}^A = \sum_{h=1}^H \tilde{n}_h^A$, $\tilde{n}^B = \sum_{l=1}^L \tilde{n}_l^B$, $W_h^A = N_h^A/N_A$, and $W_l^B = N_l^B/N_B$.

It is a common practice to sample the psu's without replacement with inclusion probabilities proportional to size. But at the stage of variance estimation, the calculations are greatly simplified by treating the sample as if the psu's were sampled with replacement. This approximation generally leads to overestimation of the variance of the estimated total, but the relative bias will not be large if the first-stage sampling fractions are not large. Jackknife variance estimators and other resampling variance estimators use this set-up. We follow the same approach for multiple frame surveys.

Denote the psu inclusion probabilities in frame A as $\tilde{\pi}_{hi}^A = \tilde{n}_h^A p_{hi}^A$, where p_{hi}^A is the normalized size measure with $\sum_i p_{hi}^A = 1$. Let $\mathbf{A}$ be a q-vector of population totals for frame A, and let $\mathbf{B}$ be an r-vector of population totals for frame B. Then $\mathbf{A}$ is estimated by

$$\hat{\mathbf{A}} = \sum_{h=1}^H \sum_{i=1}^{\tilde{n}_h^A} \frac{\hat{\mathbf{A}}_{hi}}{\tilde{\pi}_{hi}^A} = \sum_{h=1}^H \sum_{i=1}^{\tilde{n}_h^A} N_h^A \frac{\mathbf{a}_{hi}}{\tilde{n}_h^A} = \sum_{h=1}^H N_h^A \bar{\mathbf{a}}_h,$$

where $\hat{\mathbf{A}}_{hi}$ is an unbiased estimator of the population totals in sample psu i of stratum h, and $\mathbf{a}_{hi} = \hat{\mathbf{A}}_{hi}/(N_h^A p_{hi}^A)$. The estimators $\hat{\mathbf{B}} = \sum_{l=1}^L \sum_{j=1}^{\tilde{n}_l^B} N_l^B \mathbf{b}_{lj}/\tilde{n}_l^B$ and $\mathbf{b}_{lj}$ are defined similarly.

Under the assumption of with-replacement sampling, the $\mathbf{a}_{hi}$'s are independent unbiased estimators of the population mean in stratum h of frame A. Similarly, the $\mathbf{b}_{lj}$'s are independent unbiased estimators of the population mean in stratum l of frame B.

Asymptotic results. For the purposes of developing asymptotic theory, we consider parameters that may be expressed as functions of the population means $\bar{\mathbf{A}} = \mathbf{A}/N_A$ and $\bar{\mathbf{B}} = \mathbf{B}/N_B$. Consider a parameter of the form

$$\tau = g(\bar{\mathbf{A}}, \bar{\mathbf{B}}).$$

The population means $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ are estimated by $\hat{\bar{\mathbf{A}}} = \sum_{h=1}^H W_h^A \bar{\mathbf{a}}_h$ and $\hat{\bar{\mathbf{B}}} = \sum_{l=1}^L W_l^B \bar{\mathbf{b}}_l$, and τ is estimated by

$$\hat{\tau} = g(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}}).$$

Estimators of population totals discussed in Section 2 may be studied in this framework by writing $Y = N\bar{Y} = Ng(\bar{\mathbf{A}}, \bar{\mathbf{B}})$. If, for example, $\mathbf{A} = (Y_a, Y_{ab}, N_{ab}, N_A)^T$, $\mathbf{B} = (Y_b, Y_{ab}, N_{ab}, N_B)^T$, and

$$g(\bar{\mathbf{A}}, \bar{\mathbf{B}}) = \frac{N_A}{N}(\bar{A}_1 + \theta \bar{A}_2) + \frac{N_B}{N}[\bar{B}_1 + (1-\theta)\bar{B}_2],$$

then $Ng(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}}) = \hat{Y}_H(\theta)$. The PML estimator, for fixed θ, may be expressed in similar fashion as a function of $\hat{\bar{\mathbf{A}}}$ and $\hat{\bar{\mathbf{B}}}$ by noting that $\hat{N}_{ab}^{PML}(\theta)$ is a function of $\hat{\bar{A}}_3$ and $\hat{\bar{B}}_3$, and then using (1).

Define

$$\mathbf{S}_h^A = (\tilde{n}_h^A - 1)^{-1} \sum_{i=1}^{\tilde{n}_h^A} (\mathbf{a}_{hi} - \bar{\mathbf{a}}_h)(\mathbf{a}_{hi} - \bar{\mathbf{a}}_h)^T$$

and

$$\mathbf{S}_l^B = (\tilde{n}_l^B - 1)^{-1} \sum_{j=1}^{\tilde{n}_l^B} (\mathbf{b}_{lj} - \bar{\mathbf{b}}_l)(\mathbf{b}_{lj} - \bar{\mathbf{b}}_l)^T.$$

Then $\mathbf{S}_h^A$ estimates the variance of $\sqrt{\tilde{n}_h^A}\bar{\mathbf{a}}_h$, say $\boldsymbol{\Sigma}_h^A$, and $\mathbf{S}^A = \sum_{h=1}^H (W_h^A)^2 \mathbf{S}_h^A/\tilde{n}_h^A$ estimates $\boldsymbol{\Sigma}^A = \sum_{h=1}^H (W_h^A)^2 \boldsymbol{\Sigma}_h^A/\tilde{n}_h^A$, the variance of $\hat{\bar{\mathbf{A}}}$. Similarly, the variance of $\hat{\bar{\mathbf{B}}}$, say $\boldsymbol{\Sigma}^B$, is estimated by $\mathbf{S}^B = \sum_{l=1}^L (W_l^B)^2 \mathbf{S}_l^B/\tilde{n}_l^B$.

Using the asymptotic setup of Isaki and Fuller (1982), we need the following conditions. Conditions (A1) and (A2) were used in Rao and Wu (1985) to investigate properties of variance estimators in multistage stratified samples; condition (A3) ensures that the sample from one frame does not dominate the other sample asymptotically.

A1. $W_h^A \tilde{n}^A/\tilde{n}_h^A = O(1)$ and $W_l^B \tilde{n}^B/\tilde{n}_l^B = O(1)$ for all h and l. In addition, assume that $\sum_h W_h^A \boldsymbol{\Sigma}_h^A = O(1)$ and $\sum_l W_l^B \boldsymbol{\Sigma}_l^B = O(1)$.

A2. Let $g_A(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})$ be the q-vector of first derivatives with respect to the components of $\mathbf{a}$, evaluated at $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$. Analogously, $g_B(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})$ is the r-vector of first derivatives with respect to the components of $\mathbf{b}$, evaluated at $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$. Similarly, let $g''_A(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})$ be the $q \times q$ matrix of second derivatives $\partial^2 g/\partial a_j \partial a_k$, and $g''_B(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})$ be the $r \times r$ matrix of second derivatives $\partial^2 g/\partial b_j \partial b_k$, evaluated at $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$. Assume that g''_A and g''_B are continuous and bounded in a neighborhood of $(\bar{\mathbf{A}}, \bar{\mathbf{B}})$.

A3. $\tilde{n}^A/\tilde{n} \to k \in (0,1)$, where $\tilde{n} = \tilde{n}^A + \tilde{n}^B$.

Theorem 1. Under conditions (A1) to (A3),

$$\begin{aligned} V(\hat{\tau}) = \; & g_A^T(\bar{\mathbf{A}}, \bar{\mathbf{B}}) \boldsymbol{\Sigma}^A g_A(\bar{\mathbf{A}}, \bar{\mathbf{B}}) \\ & + g_B^T(\bar{\mathbf{A}}, \bar{\mathbf{B}}) \boldsymbol{\Sigma}^B g_B(\bar{\mathbf{A}}, \bar{\mathbf{B}}) + o(\tilde{n}^{-1}). \end{aligned}$$

In addition, the linearization variance estimator

$$\begin{aligned} v_L(\hat{\tau}) &= g_A^T(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}}) \mathbf{S}^A g_A(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}}) \\ &\quad + g_B^T(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}}) \mathbf{S}^B g_B(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}}) \\ &= V(\hat{\tau}) + o_p(\tilde{n}^{-1}). \end{aligned}$$

Conditions (A1) through (A3) are also used to show the consistency of the jackknife variance estimator, $v_J(\hat{\tau})$, by showing its asymptotic equivalence to the linearization variance estimator, $v_L(\hat{\tau})$.

Let $\hat{\tau}^A_{(hi)}$ be the estimator of the same form as $\hat{\tau}$ when the observations of sample psu i of stratum h are omitted:

$$\hat{\tau}^A_{(hi)} = g(\hat{\bar{\mathbf{A}}}_{(hi)}, \hat{\bar{\mathbf{B}}}),$$

where $\hat{\bar{\mathbf{A}}}_{(hi)}$ is the estimator of $\bar{\mathbf{A}}$ computed after omitting sample psu i of stratum h in frame A. Similarly, let

$$\hat{\tau}^B_{(lj)} = g(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}}_{(lj)}),$$

where $\hat{\bar{\mathbf{B}}}_{(lj)}$ is the estimator of $\bar{\mathbf{B}}$ computed after omitting sample psu j of stratum l in frame B. A jackknife variance estimator of $\hat{\tau}$ is then given by

$$\begin{aligned} v_J(\hat{\tau}) = \; & \sum_{h=1}^H \frac{\tilde{n}_h^A - 1}{\tilde{n}_h^A} \sum_{i=1}^{\tilde{n}_h^A} (\hat{\tau}^A_{(hi)} - \hat{\tau})^2 \\ & + \sum_{l=1}^L \frac{\tilde{n}_l^B - 1}{\tilde{n}_l^B} \sum_{j=1}^{\tilde{n}_l^B} (\hat{\tau}^B_{(lj)} - \hat{\tau})^2. \end{aligned} \tag{3}$$

Theorem 2. Suppose that conditions (A1) - (A3) hold. Then

$$v_J(\hat{\tau}) = v_L(\hat{\tau}) + o_p(\tilde{n}^{-1}). \tag{4}$$

The single frame estimator is expressible as a smooth function of population means. The other estimators are also smooth functions of population means, as long as the parameters θ_P, θ_H, and β_{FB} are fixed and not estimated from the data. Thus Theorem 2 shows that the jackknife variance estimator is consistent for the optimal form of each estimator.

Full and modified jackknife. The estimators $\hat{\theta}_P$, $\hat{\theta}_H$, and $\hat{\beta}_{FB}$, however, are functions of $\mathbf{S}^A$ and $\mathbf{S}^B$, which cannot be expressed as differentiable functions of means in general stratified samples. Thus Theorem 2 does not always apply directly to estimators that are functions of $\mathbf{S}^A$ and $\mathbf{S}^B$. An additional difficulty in using the jackknife can occur in highly stratified samples, because

$$\begin{aligned} \mathbf{S}^A_{(hi)} = \; & \mathbf{S}^A + \frac{(W_h^A)^2}{\tilde{n}_h^A - 2} \left[\frac{2\mathbf{S}_h^A}{\tilde{n}_h^A} \right. \\ & \left. - \frac{\tilde{n}_h^A}{(\tilde{n}_h^A - 1)^2} (\mathbf{a}_{hi} - \bar{\mathbf{a}}_h)(\mathbf{a}_{hi} - \bar{\mathbf{a}}_h)^T \right] \end{aligned} \tag{5}$$

cannot be calculated when $\tilde{n}_h^A = 2$. Similarly, $\mathbf{S}^B_{(lj)}$ cannot be calculated when $\tilde{n}_l^B = 2$. We provide a modified jackknife to handle the case of $\tilde{n}_h^A = 2$ or $\tilde{n}_l^B = 2$.

The jackknife (or a modification of the jackknife for two-psu-per-stratum designs which we shall introduce below) still provides a consistent estimate of the variance when the estimator depends on $\mathbf{S}^A$ and $\mathbf{S}^B$. Suppose

$$\hat{\tau} = g(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}}) = f(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}}, \beta),$$

with

$$\beta = [\mathbf{E}^A + \mathbf{E}^B]^{-1} [\mathbf{e}^A + \mathbf{e}^B],$$

where each element of the $p \times p$ matrix $\mathbf{E}^A$ is a linear combination of elements of $\boldsymbol{\Sigma}^A$, each element of the p-vector $\mathbf{e}^A$ is a linear combination of elements of $\boldsymbol{\Sigma}^A$, and analogously for $\mathbf{E}^B$ and $\mathbf{e}^B$. Then the estimators $\hat{Y}_{PML}(\hat{\theta}_P)$, $\hat{Y}_H(\hat{\theta}_H)$, and $\hat{Y}_{FB}(\hat{\beta}_{FB})$ are of the form $N\hat{\zeta} = Nf(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}}, \hat{\beta})$, where $\hat{\beta}$ substitutes $\mathbf{S}^A$ and $\mathbf{S}^B$ for the population covariances in β. For example, if $\mathbf{A} = (Y_a, Y_{ab}, N_{ab}, N_A)^T$ and $\mathbf{B} = (Y_b, Y_{ab}, N_{ab}, N_B)^T$, then Hartley's estimator $\hat{Y}_H(\hat{\beta}_H)$ may be written as

$$\hat{Y}_H(\beta_H) + (\hat{\beta}_H - \beta_H)[N_A(\hat{\bar{A}}_2 - \bar{A}_2) - N_B(\hat{\bar{B}}_2 - \bar{B}_2)],$$

where

$$\beta_H = -\frac{\kappa^2 \boldsymbol{\Sigma}^A(1,2) - \boldsymbol{\Sigma}^B(1,2) - \boldsymbol{\Sigma}^B(2,2)}{\kappa^2 \boldsymbol{\Sigma}^A(2,2) + \boldsymbol{\Sigma}^B(2,2)},$$

$\kappa = N_A/N_B$, and $\boldsymbol{\Sigma}^A(1,2)$ denotes the (1,2) element of $\boldsymbol{\Sigma}^A$. For the PML method, the appropriate parameter is $\beta_P = \kappa^2 \boldsymbol{\Sigma}^A(3,3)/\boldsymbol{\Sigma}^B(3,3)$.

Let $\hat{\zeta}^A_{(hi)}$ be the estimator of the same form as $\hat{\zeta}$ when the observations of sample psu i of stratum h are omitted:

$$\hat{\zeta}^A_{(hi)} = f(\hat{\bar{\mathbf{A}}}_{(hi)}, \hat{\bar{\mathbf{B}}}, \hat{\beta}^A_{(hi)}),$$

where $\hat{\beta}^A_{(hi)}$ is the estimator of β using $\mathbf{S}^A_{(hi)}$ and $\mathbf{S}^B$ (assuming $\tilde{n}^A_h > 2$). Similarly,

$$\hat{\zeta}^B_{(lj)} = f(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}}_{(lj)}, \hat{\beta}^B_{(lj)}),$$

provided $\tilde{n}^B_l > 2$. A jackknife variance estimator of $\hat{\zeta}$ is then given by

$$\begin{aligned} v_J(\hat{\zeta}) = & \sum_{h=1}^{H} \frac{\tilde{n}^A_h - 1}{\tilde{n}^A_h} \sum_{i=1}^{\tilde{n}^A_h} (\hat{\zeta}^A_{(hi)} - \hat{\zeta})^2 \\ & + \sum_{l=1}^{L} \frac{\tilde{n}^B_l - 1}{\tilde{n}^B_l} \sum_{j=1}^{\tilde{n}^B_l} (\hat{\zeta}^B_{(lj)} - \hat{\zeta})^2. \end{aligned}$$

A modified jackknife variance estimator, $v_{MJ}(\hat{\zeta})$, has the same form as $v_J(\hat{\zeta})$ but uses $\hat{\beta}$ rather than $\hat{\beta}^A_{(hi)}$ or $\hat{\beta}^B_{(lj)}$ when calculating $\hat{\zeta}^A_{(hi)}$ or $\hat{\zeta}^B_{(lj)}$. The modified jackknife variance estimator may be used when $\tilde{n}^A_h = 2$ or $\tilde{n}^B_l = 2$. A Taylor linearization variance estimator is of the form

$$\begin{aligned} v_L(\hat{\zeta}) = & \hat{g}^T_A(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}})\mathbf{S}^A \hat{g}_A(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}}) \\ & + \hat{g}^T_B(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}})\mathbf{S}^B \hat{g}_B(\hat{\bar{\mathbf{A}}}, \hat{\bar{\mathbf{B}}}), \end{aligned}$$

where $\hat{g}_A$ and $\hat{g}_B$ use $\hat{\beta}$ rather than β.

The three variance estimators $v_J(\hat{\zeta})$, $v_{MJ}(\hat{\zeta})$, and $v_L(\hat{\zeta})$ are asymptotically equivalent in the sense that the difference between each pair is of order $o_p(\tilde{n}^{-1})$. Details are given in Lohr and Rao (1997).

The results in this section apply to general stratified multistage designs in which psu's are selected with replacement. The case of stratified simple random sampling in one of the frames, say a list frame B, follows as a special case, provided the sampling fractions are negligible. Here, $\mathbf{b}_{lj}$ is a vector of values associated with the j^{th} unit in stratum l of frame B, and $\tilde{n}^B_l = n^B_l$ is the number of units sampled from the N^B_l units in stratum l of frame B. If the sampling fractions n^B_l/N^B_l are not negligible, then we replace $(n^B_l - 1)/n^B_l$ by $[(n^B_l - 1)/n_B][1 - n^B_l/N^B_l]$, similar to the case of single frame jackknife (Wolter, 1985, p. 176).

Calculating jackknife variance estimates. Estimators of Y in Section 2 may be written in the form

$$\hat{Y} = \sum_{t \in S_A} \tilde{w}^A_t y_t + \sum_{t \in S_B} \tilde{w}^B_t y_t \tag{6}$$

for modified weights $\tilde{w}^A_t$ and $\tilde{w}^B_t$. For example, $\hat{Y}_{PML}(\theta)$ uses

$$\tilde{w}^A_t = \begin{cases} w^A_t[N_A - \hat{N}^{PML}_{ab}(\theta)]/\hat{N}^A_a & \text{if } t \in a \\ w^A_t \theta\, \hat{N}^{PML}_{ab}(\theta)/\hat{N}_{ab}(\theta) & \text{if } t \in ab \end{cases}$$

and

$$\tilde{w}^B_t = \begin{cases} w^B_t[N_B - \hat{N}^{PML}_{ab}(\theta)]/\hat{N}^B_b & \text{if } t \in b \\ w^B_t(1-\theta)\, \hat{N}^{PML}_{ab}(\theta)/\hat{N}_{ab}(\theta) & \text{if } t \in ab \end{cases}.$$

If θ is estimated by $\hat{\theta}_P$, we replace θ by $\hat{\theta}_P$ in the above weights $\tilde{w}^A_t$ and $\tilde{w}^B_t$.

To calculate the jackknife estimates $\hat{Y}^A_{(hi)}$ we simply replace the weights w^A_t by the jackknife weights $w^A_{t(hi)}$: If the unit t is in cluster k of stratum g and frame A, then $w^A_{t(hi)} = 0$ if $(hi) = (gk)$; $w^A_{t(hi)} = \tilde{n}^A_h/(\tilde{n}^A_h - 1)w^A_t$ if $h = g$ and $i \neq k$; and $w^A_{t(hi)} = w^A_t$ if $h \neq g$. Similarly, we obtain $\hat{Y}^B_{(lj)}$ from the corresponding jackknife weights $w^B_{t(lj)}$. Then we estimate the variance of $\hat{Y}$ using (3).

If the modified weights $\tilde{w}^A_t$ and $\tilde{w}^B_t$ depend on elements of $\mathbf{S}^A$ and $\mathbf{S}^B$, as occurs for $\hat{Y}_{PML}(\hat{\theta}_P)$, then the full jackknife requires that elements of $\mathbf{S}^A_{(hi)}$ be computed in order to calculate $\hat{Y}^A_{(hi)}$. The matrix $\mathbf{S}^A_{(hi)}$, for $\tilde{n}^A_h \geq 3$, may be calculated by applying the jackknife again, this time to the data set with observations in psu i of stratum h deleted. Alternatively, equation (5) may be used to speed calculation.

4. Simulation Results

To study empirical properties of the variance estimators, we used the simulation study design described in detail in Skinner and Rao (1996). The population was presumed infinite, and the sample design for each frame had one stratum. A two-stage cluster sample with $\tilde{n}^A$ clusters and thirty elements per cluster was generated as the sample from frame A and a simple random sample with n_B observations was generated as the sample from frame B.

We generated 10,000 datasets for each combination of the design parameters. From each dataset we calculated $\hat{Y}_{PML}(\hat{\theta}_P)$ and the three variance estimates presented in Section 3: linearization (L), full jackknife (J), and modified jackknife (MJ). Table 1 gives some of the simulation results for the variance estimators, using $N_a/N = 0.1$, $N_b/N = 0.2$, $\mu_a = 9$, $\mu_{ab} = 10$, and $\mu_b = 11$. The empirical mean squared error (the average squared deviation of the estimate from the true value) and all variances in Table 1 were multiplied by 100 to improve readability.

The simulation results in Table 1, and other simulations performed, demonstrate that all three estimators of the variance grow closer to the empirical MSE as the sample sizes increase. For the smaller sample sizes, though, the linearization and the modified jackknife methods substantially underestimate

Table 1: Simulation results for variance estimators.

$\tilde{n}^A$	n_B	avar	EM	L	J	MJ
10	100	4.61	4.92	4.39	5.03	4.40
				(1.61)	(2.69)	(1.60)
20	100	2.79	2.75	2.75	2.91	2.77
				(.80)	(1.01)	(.81)
20	200	2.30	2.39	2.25	2.41	2.25
				(.58)	(.77)	(.58)

NOTE: avar is the theoretical asymptotic variance of the estimate, and EM is the Monte Carlo mean squared error for the 10,000 simulation runs. L, J, and MJ represent the averages of the 10,000 variance estimates for the linearization, full jackknife, and modified jackknife methods, respectively. Standard deviations are in parentheses.

the empirical MSE because they do not account for the extra variability incurred by estimating θ_P from the data. The full jackknife does not share this negative bias.

The full jackknife, though, is less stable than the other two estimators of the variance. When $\tilde{n}^A = 10$, the sample standard deviation of the full jackknife estimator of the variance is much higher than that of the linearization estimator. However, the stability improves as the sample sizes increase.

5. Discussion

The jackknife estimator of the variance has been theoretically justified, and has exhibited smaller bias than the linearization estimator of the variance in a simulation study. An advantage of the jackknife is that it is readily applied to nonlinear functions such as the ratio of two population totals. The partial derivatives used in linearization variance estimators of such nonlinear quantities are more complicated in dual frame surveys than in single frame surveys; these calculations can be avoided altogether through using the jackknife.

Other methods of variance estimation that are commonly used include balanced repeated replication (BRR) and the bootstrap, and dual frame variance estimators may be developed for these methods along the lines of the jackknife variance estimators by using appropriate weights. The advantage of BRR and bootstrap is that they can be applied to nonsmooth functions. However, for estimators $\hat{Y}$ that depend on $\mathbf{S}^A$ or $\mathbf{S}^B$, BRR and bootstrap methods also need modification in a two-psu-per-stratum design, as was noted in Section 3.

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.

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

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

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

Isaki, C. T., and Fuller, W. A. (1982), "Survey Design Under the Regression Superpopulation Model," *Journal of the American Statistical Association*, 77, 89–96.

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. (1997), "Jackknife Variance Estimation in Dual Frame Surveys," Technical Report, Department of Mathematics and Statistics, Carleton University.

Lund, R. E. (1968), "Estimators in Multiple Frame Surveys," in *ASA Proceedings of the Social Statistics Section*, 282–288.

Rao, J. N. K., and Wu, C. F. J. (1985), "Inference from Stratified Samples: Second-Order Analysis of Three Methods for Nonlinear Statistics," *Journal of the American Statistical Association*, 80, 620-630.

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

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

Vogel, F. A. (1975), "Surveys with Overlapping Frames—Problems in Applications," *ASA Proceedings of the Social Statistics Section*, 694-699.

Wolter, K. M. (1985), *Introduction to Variance Estimation*, New York: Springer-Verlag.