A Resampling Procedure for Complex Survey Data

R. R. SITTER*

In complex survey data, often the sampling design induces a non-iid structure to the data (e.g., without replacement sampling, stratification, multistage, or unequal probability of selection). Though techniques for variance estimation and confidence intervals do exist, they often are cumbersome to implement or do not extend to complex designs. It would be desirable to have resampling methods that reuse the existing estimation system repeatedly, using computing power to avoid theoretical work, and that can be applied to such data. In recognition of this need, various resampling procedures for variance estimation and confidence intervals in sample survey data (where the sampling is without replacement) have been proposed in the literature. These include the jackknife, the with-replacement bootstrap (BWR), the without-replacement bootstrap (BWO), and the rescaling bootstrap. The BWR and BWO are applicable only to simple sampling designs. Others have shown the asymptotic consistency of jackknife variance estimates for nonlinear functions of means for multistage designs in which the primary sampling units are selected with replacement. However, when the primary sampling units are selected without replacement, the jackknife has been developed only for stratified sampling. The rescaling bootstrap extends to more complex sampling designs, but is applicable only to functions of means and can be computationally more intensive and difficult to use. The resampling method developed in this article retains the desirable properties of the BWR and BWO, but extends to more complex without-replacement sampling designs: (a) stratified random sampling, (b) two-stage cluster sampling, and (c) the Rao–Hartley–Cochran method of unequal probability sampling. The method consists of resampling without replacement from the data vector to mirror the original sampling design and then repeating this with replacement to match the usual variance estimates in the linear case. In the simple sampling designs to which the BWR is applicable, this resampling method contains the BWR as a special case and consequently suggests an extension of the BWR to two-stage cluster sampling. The properties of the variance estimators of $\hat{\theta}$ and the bootstrap-t confidence intervals for θ are studied. The variance estimator is shown to be consistent when $\hat{\theta}$ is a nonlinear function of means, in situation (a). The confidence intervals are shown to capture the second-order term of the Edgeworth expansion of the distribution of $\bar{y}$ in situation (a) and thus account for skewness in the distribution. A simulation study for situation (a) suggests that the confidence intervals from the method track the nominal one-tailed error rates better than do the normal theory intervals, but the variance estimates are less stable. For the median, the method generally performs very well.

KEY WORDS: Bootstrap; Edgeworth expansion; Jackknife; Probability sampling; Rao–Hartley–Cochran method; Two-stage cluster sampling.

Although techniques for variance estimation and confidence intervals exist for complex survey data, they often are cumbersome to implement or do not extend to complex designs or nonlinear estimators. The three most commonly used methods are the linearization (or Taylor) method, the jackknife method, and balanced repeated replications (BRR). Krewski and Rao (1981) showed the asymptotic consistency of the variance estimates for nonlinear functions of means based on these three methods applied to multistage designs in which the primary sampling units are selected with replacement. These methods all have certain drawbacks, however; the linearization method requires theoretical calculation and subsequent programming of derivatives, which can make it cumbersome to implement, and the BRR and the jackknife methods have been developed only for stratified sampling when the primary sampling units are selected without replacement. It would be desirable to have resampling methods that resuse the existing estimation system repeatedly, using computing power to avoid theoretical work, and that can be applied to complex survey designs. In recognition of this need, various resampling procedures for variance estimation and confidence intervals in sample survey data have been proposed in the literature (McCarthy and Snowden 1985; Rao and Wu 1988).

This article aims to explore extensions of the bootstrap to complex survey data, where the sampling is without replacement. In Section 1, a resampling method for stratified simple random sampling without replacement is proposed. The properties of the resulting variance estimator and an estimate of bias are explored. Further, with appropriate choice of resample size, the method is shown to match the usual unbiased estimate of the third moment of the sample mean. Also provided are a further theoretical justification for this choice of resample size based on Edgeworth expansions, some suggestions on randomization to handle noninteger resample sizes and a partial summary of a larger simulation study by Sitter (1990). In Section 2 the method is extended to two-stage cluster sampling without replacement. In Section 3 the method is extended to the Rao–Hartley–Cochran (1962) method of unequal probability sampling.

1. STRATIFIED RANDOM SAMPLING

In stratified random sampling the finite population, consisting of N units, is partitioned into L nonoverlapping strata of $N_1, N_2, \ldots, N_L$ units; thus, $N_1 + N_2 + \cdots + N_L = N$. A simple random sample without replacement (SRSWOR) is taken independently from each stratum. The sample sizes within each stratum are denoted by $n_1, n_2, \ldots, n_L$, and the total sample size is $n = n_1 + n_2 + \cdots + n_L$. A vector of measurements of some unit characteristics is represented as $\mathbf{y}_{hi} = (y_{1hi}, y_{2hi}, \ldots, y_{\tau hi})^T$, where the subscript h refers to the stratum label and the subscript i refers to the ith unit

* R. R. Sitter is Assistant Professor, Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario, K1S 5B6, Canada. This research composed a portion of the author's Ph.D. thesis at the University of Waterloo, Ontario, Canada, under the supervision of C. F. J. Wu.

Journal of the American Statistical Association
September 1992, Vol. 87, No. 419, Theory and Methods

within the hth stratum. The population parameter of interest $\theta = \theta(\mathbf{S})$, where $\mathbf{S} = \{\mathbf{y}_{hi} : h = 1, 2, \ldots, L; i = 1, 2, \ldots, N_h\}$, usually is estimated by $\hat{\theta} = \hat{\theta}(\mathbf{s})$, where $\mathbf{s} = \{\mathbf{y}_{hi} : h = 1, 2, \ldots, L; i = 1, 2, \ldots, n_h\}$. The population mean vector is denoted by $\bar{\mathbf{Y}} = (\bar{Y}_1, \ldots, \bar{Y}_\tau)^T$. In this case, its unbiased estimate is $\bar{\mathbf{y}} = \sum_{h=1}^{L} W_h \bar{\mathbf{y}}_h = (\bar{y}_1, \ldots, \bar{y}_\tau)^T$, where $\bar{\mathbf{y}}_h = \sum_{i=1}^{n_h} \mathbf{y}_{hi}/n_h = (\bar{y}_{1h}, \bar{y}_{2h}, \ldots, \bar{y}_{\tau h})^T$ and $W_h = N_n/N$. For $\tau = 1$, an unbiased estimate of $\operatorname{Var}(\bar{y})$, given in Cochran (1977), is

$$\operatorname{var}(\bar{y}) = \sum_{h=1}^{L} W_h^2 \frac{1 - f_h}{n_h} s_h^2, \qquad (1.1)$$

where $f_h = n_h/N_h$ and $s_h^2 = \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2/(n_h - 1)$.

1.1 Existing Bootstrap Methods

If the standard iid bootstrap is applied to the sample data $\{\mathbf{y}_{hi}\}_{i=1}^{n_h}$ in each stratum, then the resulting resampling algorithm would take the following form:

1. Draw a simple random sample with replacement (SRSWR) $\{\mathbf{y}_{hi}^*\}_{i=1}^{n_h}$ from the original sample $\{\mathbf{y}_{hi}\}_{i=1}^{n_h}$ independently for each stratum to get $\mathbf{s}^* = \{\mathbf{y}_{hi}^* : h = 1, 2, \ldots, L; i = 1, 2, \ldots, n_h\}$. Let $\hat{\theta}^* = \hat{\theta}(\mathbf{s}^*)$.
2. Repeat step 1 a large number of times, B, to get $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$.
3. Estimate $\operatorname{var}(\hat{\theta})$ with

$$v_b = E_*(\hat{\theta}^* - E_*\hat{\theta}^*)^2$$

or its Monte Carlo approximation $\tilde{v}_b = \sum_{i=1}^{B} (\hat{\theta}_i^* - \hat{\theta}_{(\cdot)}^*)^2/B$, where $\hat{\theta}_{(\cdot)}^* = \sum_{i=1}^{B} \hat{\theta}_i^*/B$ and E_* refers to the expectation with respect to the resampling. In this equation, $E_*\hat{\theta}^*$ and $\hat{\theta}_{(\cdot)}^*$ can be replaced by $\hat{\theta}$.

In the simplest case with $\hat{\theta} = \bar{y}$,

$$v_b = \sum_{h=1}^{L} \frac{W_h^2}{n_h} \left(\frac{n_h - 1}{n_h}\right) s_h^2.$$

Comparing this with the standard unbiased variance estimate, $\operatorname{var}(\bar{y})$, given in equation (1.1), it is seen that v_b is not unbiased for $\operatorname{var}(\bar{y})$ and in the important special case of bounded n_h is not a consistent estimator of $\operatorname{var}(\bar{y})$ as $L \to \infty$. This can be avoided by using a correction factor only if $n_h = n_0$ and $f_h = f$ for all h, in which case $[n_0/(n_0 - 1)](1 - f)v_b$ is consistent and unbiased.

Two adjustments to this resampling algorithm have been proposed in the literature as solutions to this problem. McCarthy and Snowden (1985) proposed what they termed the with-replacement bootstrap (BWR). In this method, one chooses the resample size in the previously stated bootstrap algorithm to match the usual unbiased variance estimator for $\bar{y}$, $\operatorname{var}(\bar{y})$. If the recommended choice is noninteger, then a randomization between bracketing integers is available. Our method includes the BWR as a special case, which may not be the best application. Rao and Wu (1988) proposed a rescaling of the standard bootstrap when $\hat{\theta} = g(\bar{\mathbf{y}})$, a nonlinear function of means. In this method one applies the previously stated algorithm with a general resample size m_h not necessarily equal to n_h, but rescales the resampled values appropriately so that the resulting variance estimate is the same as the usual unbiased variance estimate in the linear case. Rao and Wu showed that the method yields consistent bias and variance estimates and, with appropriate choice of the resample size, the resulting histogram matches the second-order term of the Edgeworth expansion of $\bar{y}$, as $L \to \infty$. As proposed, however, this method requires rescaling each data point at each bootstrap iteration. These rescaling factors require various summary statistics of subsets of the sample. In large complex surveys this could be very undesirable.

In addition to these two with-replacement resampling schemes, a without-replacement bootstrap (BWO) was proposed by Gross (1980) in the case of a single stratum. Suppressing the h subscript, this method assumes $N = kn$ for some integer k and creates a pseudopopulation of size N by replicating the data k times. The resample is then obtained by drawing n units without replacement from the pseudopopulation. Although the BWO method is intuitively appealing, it does not yield the usual unbiased estimate of variance in the linear case. Bickel and Freedman (1984) proposed a randomization between two pseudopopulations that corrects this problem and allows an extension to $L > 1$; however, their method is applicable only for some stratified samples (see McCarthy and Snowden 1985).

1.2 The Proposed Method

It seems reasonable to design a resampling scheme that parallels the original sampling scheme as closely as possible. This is what is so appealing about the BWO method as compared to the BWR method and the rescaling method. The BWO method resamples without replacement with the same sampling fraction (within each stratum) as the original sampling plan. In contrast, the BWR and rescaling methods, are more in the spirit of the standard bootstrap. Unfortunately, the BWO method does not yield reasonable estimates except in the simplest sampling plans. In general, the method we propose entails: (a) selecting a subsample without replacement from the original sample to mirror the original sampling scheme; (b) replacing this subsample; and (c) repeating this a specified number of times, k_h. This bootstrapping procedure is repeated a large number of times. The value of k_h is chosen such that the bootstrap estimate of variance matches the usual one in the linear case. With these points in mind, consider the following procedure: Assume $N_h = n_h k_h$ ($\Leftrightarrow k_h = f_h^{-1}$) and $n_h' = f_h n_h$, with both k_h and n_h' integers greater than or equal to 1, for each $h = 1, 2, \ldots, L$. (The integer assumption can be handled using a randomization; see Section 1.6. The assumption $n_h' \geq 1$, however, is restrictive.) Then follow these steps:

1. Resample n_h' without replacement from stratum h to get $\mathbf{y}_{h1}^*, \mathbf{y}_{h2}^*, \ldots, \mathbf{y}_{hn_h'}^*$.
2. Repeat step 1 k_h times independently, replacing the resamples of size n_h' each time, to get $\mathbf{y}_{h1}^*, \mathbf{y}_{h2}^*, \ldots, \mathbf{y}_{hn_h}^*$.
3. Repeat steps 1 and 2 independently for each stratum

to get $\mathbf{s}^* = \{\mathbf{y}^*_{hi} : h = 1, 2, \ldots, L; i = 1, 2, \ldots, n_h\}$, and let $\hat{\theta}^* = \hat{\theta}(\mathbf{s}^*)$.

4. Repeat steps 1–3 a large number of times, B, to get $\hat{\theta}^*_1, \hat{\theta}^*_2, \ldots, \hat{\theta}^*_B$.

5. Estimate $\mathrm{var}(\hat{\theta})$ with

$$v = E_*(\hat{\theta}^* - E_*\hat{\theta}^*)^2$$

or its Monte Carlo approximation $\tilde{v} = \sum_{i=1}^{B} (\hat{\theta}^*_i - \hat{\theta}^*_{(\cdot)})^2/B$. Note that $E_*\hat{\theta}^*$, and $\hat{\theta}^*_{(\cdot)}$ can be replaced by $\hat{\theta}$.

Note that in the without-replacement step, the resampling fraction is the same as the original sampling fraction. This is then repeated k_h times independently to obtain a resample vector of size n_h.

Under the stated assumptions if $\theta = \bar{Y}$ and $\hat{\theta} = \bar{y}$, then $v = \mathrm{var}(\bar{y})$ [see equation (1.2)], the usual unbiased estimate of $\mathrm{Var}(\bar{y})$, and $E_*(\bar{y}^* - \bar{y})^3 = \hat{\mu}_3(\bar{y})$, where $m_{3h} = \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^3/n_h$, the sample third moment of stratum h, and

$$\hat{\mu}_3(\bar{y}) = \sum_{1}^{L} W_h^3 \frac{(1 - f_h)(1 - 2f_h)m_{3h}}{(n_h - 1)(n_h - 2)}$$

is the usual unbiased estimate of $E(\bar{y} - \bar{Y})^3$; see Section 1.3. Also note that if $\hat{\theta} = g(\bar{\mathbf{y}})$, then $\hat{\theta}^* = g(\bar{\mathbf{y}}^*)$. What one effectively has done is to avoid the rescaling by adjusting the resampling scheme to mirror the original sampling plan. Also if $\hat{\theta}$ is nonnegative by definition, then so is $\hat{\theta}^*$. This property is not shared by the rescaling method (see Rao and Wu 1988).

The problem with the previously outlined procedure is the restriction $n'_h = f_h n_h \geq 1$. To illustrate the reason for such a restriction, consider a single stratum and suppress subscript h. If $n' = fn \geq 1$ but is not integer valued, then one could choose the resample size $n'_1 = \lfloor fn \rfloor$ with probability p and $n'_2 = \lceil fn \rceil$ with probability $1 - p$, where p is chosen so that the resulting estimate of variance of $\bar{y}$ is the usual unbiased estimate ($\lfloor \cdot \rfloor$ denotes the greatest integer less than, and $\lceil \cdot \rceil$ denotes the smallest integer greater than). If, however, $0 < n' = fn < 1$, then this randomization is not sensible because the resample size should be at least 1. With this in mind the following method is proposed:

1. Choose $1 \leq n'_h < n_h$ and resample n'_h units without replacement from stratum h to get $\mathbf{y}^*_{h1}, \mathbf{y}^*_{h2}, \ldots, \mathbf{y}^*_{hn'_h}$.

2. Repeat step 1 $k_h = [n_h(1 - f^*_h)]/[n'_h(1 - f_h)]$ times independently, replacing the resamples of size n'_h each time, to get $\mathbf{y}^*_{h1}, \mathbf{y}^*_{h2}, \ldots, \mathbf{y}^*_{hn^*_h}$ (for k_h noninteger, see Section 1.6), where $f^*_h = n'_h/n_h$ and $n^*_h = k_n n'_h$.

3. Do steps 3, 4, and 5 as described previously on this page.

In the linear case with $\tau = 1$, $\theta = \bar{Y}$, and $\hat{\theta} = \bar{y}$, it is clear that $E_*(\bar{y}^*) = \bar{y}$; thus v becomes

$$v = V_*(\bar{y}^*) = \sum_{h=1}^{L} W_h^2 V_*(\bar{y}^*_h),$$

where we can write $\bar{y}^*_h = \sum_{j=1}^{k_h} \sum_{l=1}^{n'_h} y^*_{hjl}/k_h n'_h$, and the j and l subscripts refer to units resampled in the with-replacement and the without-replacement steps. So

$$\begin{aligned} V_*(\bar{y}^*) &= \sum_{h=1}^{L} W_h^2 V_*\left(\frac{1}{k_h n'_h} \sum_{j=1}^{k_h} \sum_{l=1}^{n'_h} y^*_{hjl}\right) \\ &= \sum_{h=1}^{L} \frac{W_h^2}{k_h} V_*\left(\frac{1}{n'_h} \sum_{l=1}^{n'_h} y^*_{h1l}\right) \\ &= \sum_{h=1}^{L} W_h^2 \frac{(1 - f^*_h)}{k_h n'_h} s_h^2. \end{aligned} \tag{1.2}$$

Using the fact that $f^*_h = n'_h/n_h$ and $k_h = [n_h(1 - f^*_h)]/[n'_h(1 - f_h)]$, this reduces to the usual unbiased variance estimate, $\mathrm{var}(\bar{y})$. This holds for any choice of n'_h under the integer assumptions of the procedure. The intuitive choice for the n'_h's are the ones that match the resampling fraction to the original sampling fraction, as was done in the first algorithm. Theoretical justification for this choice is given in Sections 1.3 and 1.5.

Appendix A shows that in the nonlinear case, with $\hat{\theta} = g(\bar{\mathbf{y}})$,

$$v = v_L + O_p(n^{-2}) \tag{1.3}$$

under the conditions stated therein, where

$$v_L = \sum_{\alpha,\beta=1}^{\tau} g_\alpha(\bar{\mathbf{y}}) g_\beta(\bar{\mathbf{y}}) \sum_{h=1}^{L} W_h^2 \frac{(1 - f_h)}{n_h} s_{\alpha\beta h},$$

$g_\alpha(\mathbf{z}) = \partial g(\mathbf{z})/\partial z_\alpha$, $\mathbf{z} = (z_1, \ldots, z_\tau)^T$, and $(n_h - 1)s_{\alpha\beta h} = \sum_i (y_{\alpha hi} - \bar{y}_{\alpha h})(y_{\beta hi} - \bar{y}_{\beta h})$. Recall that $\mathbf{y}_{hi} = (y_{1hi}, \ldots, y_{\tau hi})^T$. Thus v is a consistent estimator of the variance of $\hat{\theta}$, because v_L (which is a variance estimator under the linearization approach) is consistent (Bickel and Freedman 1984).

One also can use the procedure to obtain a consistent estimate of the bias in the nonlinear case. Rao and Wu (1985) gave an asymptotic approximation for the bias of $\hat{\theta}$ for with-replacement sampling. Their result can be extended to the without-replacement case, under (A.1) of Appendix A, to get

$$\begin{aligned} B(\hat{\theta}) &= E(\hat{\theta}) - \theta \\ &= \frac{1}{2} \sum_{\alpha,\beta=1}^{\tau} g_{\alpha\beta}(\bar{\mathbf{Y}}) \sum_{h=1}^{L} W_h^2 \frac{(1 - f_h)}{n_h} S_{\alpha\beta h} + O_p(n^{-3/2}), \end{aligned}$$

where $g_{\alpha\beta}(\mathbf{z})$ is $\partial^2 g(\mathbf{z})/\partial z_\alpha \partial z_\beta$ and $S_{\alpha\beta h} = E(y_{\alpha hi} - \bar{Y}_{\alpha h}) \times (y_{\beta hi} - \bar{Y}_{\beta h})$. An estimate of $B(\hat{\theta})$ is $\hat{B}(\hat{\theta}) = E_*(\hat{\theta}^*) - \hat{\theta}$, which can be approximated by $\hat{\theta}^*_{(\cdot)} - \hat{\theta}$. Let $g'(\bar{\mathbf{y}}) = [g'_1(\bar{\mathbf{y}}), \ldots, g'_\tau(\bar{\mathbf{y}})]^T$ be the vector of first derivatives $\partial g(\mathbf{x})/\partial x_\alpha$ evaluated at $\mathbf{x} = \bar{\mathbf{y}}$. Let $g''(\bar{\mathbf{y}}) = \{g_{\alpha\beta}(\bar{\mathbf{y}})\}$ be the $\tau \times \tau$ matrix of second derivatives $\partial g^2(\mathbf{x})/\partial x_\alpha \partial x_\beta$ evaluated at $\mathbf{x} = \bar{\mathbf{y}}$. Using (A.1) and (A.2) of Appendix A, we get

$$\begin{aligned} \hat{B}(\hat{\theta}) &= E_*(\hat{\theta}^*) - \hat{\theta} \\ &= \frac{1}{2} \sum_{\alpha,\beta=1}^{\tau} g_{\alpha\beta}(\bar{\mathbf{y}}) \sum_{h=1}^{L} W_h^2 \frac{(1 - f_h)}{n_h} s_{\alpha\beta h} + O_p(n^{-3/2}) \\ &= B(\hat{\theta}) + O_p(n^{-3/2}). \end{aligned}$$

Rao and Wu (1988) obtained similar results for the rescaling method when sampling with replacement.

1.3 Choosing n'_h To Match Third Moments

In this section the goal is to choose n'_h in the proposed bootstrap algorithm of the previous section so that the bootstrap third moment when $\tau = 1$, $E_*(\bar{y}^* - \bar{y})^3$; is equal to $\hat{\mu}_3(\bar{y})$. For the proposed algorithm $E_*(\bar{y}^*) = \bar{y}$ and $\mathrm{var}_*(\bar{y}^*) = \mathrm{var}(\bar{y})$; thus the first two moments of the distribution of $\bar{y}^*$ match the usual unbiased estimates of the first two moments of $\bar{y}$. In this section $E_*(\bar{y}^* - \bar{y})^3$ will be considered. Recall first that, assuming $n_h > 2$, the usual unbiased estimate of $E(\bar{y} - \bar{Y})^3$ is

$$\hat{\mu}_3(\bar{y}) = \sum_1^L W_h^3 \frac{(1-f_h)(1-2f_h)m_{3h}}{(n_h-1)(n_h-2)},$$

where $m_{3h} = \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^3/n_h$, the sample third moment of stratum h. Also, assuming that k_h and n'_h (as described in section 1.2) are integers, the total resample size within the hth stratum will be

$$n_h^* = k_h n'_h = \frac{n_h(1-f_h^*)}{(1-f_h)},$$

where $f_h^* = n'_h/n_h$ is the resampling fraction. Appendix B shows that if we let $\bar{z}^* = \bar{y}^* - \bar{y}$, then

$$E_*(\bar{z}^*)^3 = \sum_{h=1}^L \left[W_h^3 \frac{(1-f_h)(1-2f_h)m_{3h}}{(n_h-1)(n_h-2)}\right][1 - A_h], \tag{1.4}$$

where

$$A_h = \frac{(f_h^* - f_h)}{(1-f_h^*)(1-2f_h)} \tag{1.5}$$

is an error-in-matching factor that becomes 0 if the resampling fraction within each stratum is matched to the original stratum sampling fraction. So when applicable, the intuitive choice of $n'_h = f_h n_h$ (described in section 1.2) matches the third moment in the linear case.

1.4 Confidence Intervals

The simplest confidence interval estimate based on the proposed method is the percentile method. The histogram of the B estimates $\hat{\theta}_b^*$ for $b = 1, \ldots, B$ are used to form the $(1 - 2\alpha)$ confidence interval $(\hat{\theta}_L(\alpha), \hat{\theta}_U(\alpha))$, where $\hat{\theta}_L(\alpha) = \hat{\mathrm{CDF}}^{-1}(\alpha)$, $\hat{\theta}_U(\alpha) = \hat{CDF}^{-1}(1-\alpha)$, and $\hat{CDF}(z) = \#\{\hat{\theta}_b^* \le z; b = 1, \ldots, B\}/B$.

Perhaps a better method in terms of coverage probability would be the bootstrap-t method (see Abramovitch and Singh 1985). In this method the distribution of $t = (\hat{\theta} - \theta)/\sqrt{v}$ is estimated by $t^* = (\hat{\theta}^* - \hat{\theta})/\sqrt{v^*}$, its bootstrap estimate, where v^* is the bootstrap estimate obtained by reapplying the proposed method to the resampled values $\{y_{hi}^*\}$ (i.e., by replacing y_{hi} by y_{hi}^* in the proposed algorithm). This double bootstrap yields $t_1^*, \ldots, t_B^*$, resampled estimates of t. One now can obtain $(1 - 2\alpha)$-lével confidence intervals for θ by

$$\{\hat{\theta} - t_U^*\sqrt{v}, \hat{\theta} - t_L^*\sqrt{v}\},$$

where t_L^* and t_U^* are the α and $1 - \alpha$ percentiles of the obtained bootstrap histogram.

Because the bootstrap-t method takes a large amount of computation, a simpler method is to use a bootstrap-t_J instead. Let v_J be a jackknife estimate of variance (Krewski and Rao 1981) and replace v and v^* by v_J and v_J^*, the jackknife variance estimator applied to the sampled and resampled data, to yield

$$\{\hat{\theta} - t_U^*\sqrt{v_J}, \hat{\theta} - t_L^*\sqrt{v_J}\},$$

where t_L^* and t_U^* are the percentiles from the distribution of $t_J^* = (\hat{\theta}^* - \hat{\theta})/\sqrt{v_J^*}$.

1.5 Edgeworth Expansions

In Section 1.3 we showed that choosing $n'_h = f_h n_h$ implies third-moment matching. In this section we further show that the same choice of n'_h captures the second-order term of the Edgeworth expansion for known strata variances, $S_h^2 = E(y_{hi} - \bar{Y}_h)^2$. One can obtain an Edgeworth expansion for the distribution of $X = (\bar{y} - \bar{Y})/\sigma(\bar{y})$, where $\sigma^2(\bar{y}) = \sum_1^L W_h^2(1 - f_h)S_h^2/n_h$, by using the standard expansion for the independent but not identically distributed case (Feller 1966; Rao and Wu 1988):

$$F_X(x) = \Phi(x) + \sum_{h=1}^L \frac{W_h^3 E(\bar{y}_h - \bar{Y}_h)^3}{6\sigma^3(\bar{y})} \Phi^{(3)}(x) + O(L^{-1}), \tag{1.6}$$

as $L \to \infty$ with $\max_h W_h = O(L^{-1})$ and $\max_h n_h = O(1)$, where

$$E(\bar{y}_h - \bar{Y}_h)^3 = \frac{N_h^2(1-f_h)(1-2f_h)\mu_{3h}}{n_h^2(N_h-1)(N_h-2)},$$

$$\mu_{3h} = E(y_{hi} - \bar{Y}_h)^3,$$

and $\Phi(x)$ and $\Phi^{(3)}(x)$ denote the standard normal cdf and its third derivative. Because the resampling also is done independently for each stratum, the same theory can be applied to obtain an Edgeworth expansion for $X^* = (\bar{y}^* - \bar{y})/\sigma(\bar{y}^*)$ given by

$$F_{X^*}(x) = \Phi(x) + \sum_{h=1}^L \frac{W_h^3 E_*(\bar{y}_h^* - \bar{y}_h)^3}{6\sigma^3(\bar{y}^*)} \Phi^{(3)}(x) + O_p(L^{-1}), \tag{1.7}$$

where, from Section 1.3,

$$E_*(\bar{y}_h^* - \bar{y}_h)^3 = \left[\frac{(1-f_h)(1-2f_h)m_{3h}}{(n_h-1)(n_h-2)}\right][1 - A_h], \tag{1.8}$$

$$\sigma^2(\bar{y}^*) = \sum_1^L W_h^2(1-f_h)s_h^2/n_h, \tag{1.9}$$

and

$$s_h^2 = \sum_1^{n_h} (y_{hi} - \bar{y}_h)^2/(n_h - 1). \tag{1.10}$$

Substituting (1.8) and (1.9) into (1.7) and applying the strong law of large numbers, one gets

$$F_{X^*}(x) = \Phi(x) + \sum_{h=1}^{L} \frac{W_h^3 E(\bar{y}_h - \bar{Y}_h)^3}{6\sigma^3(\bar{y})} \times [1 - A_h]\Phi^{(3)}(x) + O_p(L^{-1}). \quad (1.11)$$

Computing (1.11) to (1.6), it can be seen that the second term of (1.6), of order $O(L^{-1/2})$, is capture if one of the following holds: (a) $f_h^* = f_h$ for $h = 1, \ldots, L$; (b) $f_h^* \to f_h$ for $h = 1, \ldots, L$; or (c) f_h, $f_h^* = o(1)$.

Condition (a) is equivalent to $n'_h = f_h n_h$. (The underlying asymptotic framework is a sequence of finite stratified populations subscripted by u, where $L_u \to \infty$ as $u \to \infty$. In this sense conditions (b) and (c) refer to the underlying dependence of f_h on u). So the intuitive choice of $n'_h = f_h n_h$ captures the second-order term of the Edgeworth expansion as $L \to \infty$. Rao and Wu (1988) showed a similar result for the recaling method under with-replacement sampling and also showed that the Bickel and Freedman (1984) extension of the BWO method does not have this property.

Another asymptotic framework that can be considered is when L is finite and n_h, $N_h \to \infty$. This situation is more complex and involves some technical development. Chen and Sitter (1991) derived an Edgeworth expansion for $\bar{y}^*$ under this framework and compared it to the Edgeworth expansions of $\bar{y}$ derived by Robinson (1978) and Babu and Singh (1985). They showed that the proposed method captures the second-order term up to the same multiplicative factor $A = (f^* - f)/[(1 - f^*)(1 - 2f)]$. As in the previously discussed term, this term is negligible under any of conditions (a), (b), or (c). In this asymptotic framework, however, one can ensure that one of conditions (b) or (c) hold by choosing n'_h appropriately.

1.6 Randomization to Handle Noninteger Resample Size

First, it is easy to show that if $n' \le n/(2 - n/N)$, then $k = [n(1 - n'/n)]/[n'(1 - n/N)] \ge 1$. This implies that when applying the proposed method in the case of stratified random sampling, if one chooses n'_h to be an integer such that $1 \le n'_h \le n_h/(2 - f_h)$, then $k_h = [n_h(1 - f_h^*)]/[n'_h(1 - f_h)] \ge 1$ for each stratum. If k_h is not an integer, let K_h be a random variable such that

$$P(K_h = k_{h1}) = \frac{\left(\dfrac{1}{k_h} - \dfrac{1}{k_{h2}}\right)}{\left(\dfrac{1}{k_{h1}} - \dfrac{1}{k_{h2}}\right)} = p_h,$$

and

$$P(K_h = k_{h2}) = 1 - p_h,$$

where $1 \le k_{h1} < k_h < k_{h2} < n_h$ and k_{h1} and k_{h2} are integers (usually $k_{h1} = \lfloor k_h \rfloor$ and $k_{h2} = \lceil k_h \rceil$). Now apply the algorithm as described in Section 1.2, using K_h in place of k_h. The randomization must be done independently for each stratum and repeated at each bootstrap iteration. Under this revised algorithm, if $\hat{\theta} = \bar{y}$, then $\mathrm{Var}_*(\bar{y}^*) = \mathrm{var}(\bar{y})$.

If in the proposed method one uses the suggested $n'_h = f_h n_h$, then it is possible for both n'_h and k_h to be noninteger. In this case there are a number of ways to randomize between bracketing integers so that $\mathrm{Var}_*(\bar{y}^*) = \mathrm{var}(\bar{y})$. For the remainder of this section a single stratum will be considered and the h subscript will be suppressed, because the randomization would be applied independently to each stratum. A randomization of k over bracketing integers was given previously for any integer choice of n'. Here we would like to randomize n' as well. A simple method would be to let $n'_1 = \lfloor n' \rfloor$ and $n'_2 = \lceil n' \rceil$ and use n'_1 with probability p and n'_2 with probability $1 - p$, where

$$p = n'_2 - n'.$$

Given the value of n' obtained in this randomization, k could be randomized between bracketing integers as shown previously, for fixed n'. Under this two-stage randomization,

$$E(f^*) = f$$

and

$$\mathrm{Var}_*(\bar{y}^*) = \mathrm{var}(\bar{y}).$$

This is the randomization method used in the simulation study summarized in Section 1.7.

It may be noted that if this procedure is used, then one of the braketing integers for n' could suggest a value for k less than 1. This could happen only if $\lceil n' \rceil / n > \frac{1}{2}$. In this case this randomization method could not be used.

Another method that could be considered is to let n'_1 and n'_2 be defined as previously stated and let k_1 and k_2 be defined similarly, then to use the pair (n'_1, k_1) with probability p and (n'_2, k_2) with probability $1 - p$, where

$$p = \frac{\left(\dfrac{(1 - f)}{n} - \dfrac{(1 - n'_2/n)}{k_2 n'_2}\right)}{\left(\dfrac{(1 - n'_1/n)}{k_1 n'_1} - \dfrac{(1 - n'_2/n)}{k_2 n'_2}\right)}.$$

This method will work in all cases where the algorithm is applicable, but likely will not perform quite as well as the previously described method.

1.7 A Simulation Study

The results presented in this section are taken from a simulation study done by Sitter (1989) to compare the performance of the proposed resampling method to various other methods commonly used in variance estimation and confidence interval estimation for stratified random sampling without replacement. Sitter considered a number of finite populations based on hypothetical populations. The methods considered, along with the specific notation used in Tables 1 and 2 included: (a) the linearization method (LIN); (b) the delete-1 jackknife method (JACK); (c) the proposed method with n'_h chosen to match third moments (PM1) and with $n'_h = 1$ (PM2), which reduces the method to the BWR; (d) the rescaling method proposed by Rao and Wu (1988) for without-replacement sampling, with the resample size chosen to match third moments (RS1) and with the resample size equal $n_h - 1$ (RS2), in which case the rescaling adjusts

Table 1. Monte Carlo Error Rates in the Upper and Lower Tails, Standardized Lengths, Relative Biases, and Relative Stabilities of Six Methods (ratio, corr.[a]) for Stratified Sampling Without Replacement

	Nominal one-tailed error rate						Standardized length[b]			
	5%			10%						
Method	L	U	L + U	L	U	L + U	5%	10%	% bias[c]	Stab[d]
					Ratio					
LIN	4.0	5.3	9.3	9.5	9.8	19.3	.97	.97	−4.5	.21
JACK	4.0	5.2	9.2	9.5	9.8	19.3	.97	.97	−4.6	.21
PM1	6.0	5.3	11.3	11.1	9.7	20.8	.96	.95	−4.0	.23
PM2	5.5	4.8	10.3	10.3	8.7	19.0	1.00	.99	−4.5	.23
RS1	4.6	4.6	9.2	9.8	8.8	18.6	1.00	1.00	−3.8	.23
RS2	4.8	4.7	9.5	10.3	8.9	19.2	1.00	1.00	−4.4	.23
					Correlation					
LIN	8.1	3.2	11.3	14.0	6.9	20.9	.95	.95	−6.2	.37
JACK	7.3	2.7	10.0	13.0	6.3	19.3	.98	.98	4.7	.41
PM1	5.9	4.4	10.3	11.7	8.5	20.2	1.00	.98	−2.8	.39
PM2	5.2	3.3	8.5	10.4	7.8	18.2	1.05	1.02	−3.2	.39
RS1	5.8	3.5	9.4	10.4	7.9	18.3	1.04	1.02	−2.6	.38
RS2	4.7	3.5	8.2	10.0	8.6	18.6	1.06	1.03	−3.8	.38

[a] The ratio and correlation coefficient as defined on this page.
[b] Standardized length = (length of interval estimate)/($2z_\alpha \sqrt{\text{MSE}(\hat\theta)}$).
[c] % Relative bias = (E(var. est.)/MSE ($\hat\theta$) − 1) × 100%.
[d] Relative stability = $\sqrt{\text{MSE (var. est.)}}$/MSE ($\hat\theta$).

only for without-replacement sampling; and (e) a method based on Woodruff's (1952) confidence intervals for quantiles (WOOD). Woodruff's method involves inverting the standard estimate of the cdf (see Kovar, Rao, and Wu 1988 or Sitter 1992).

The parameters θ considered were the ratio, the correlation coefficient, and the median. Assuming the notation for stratified sampling of Section 1, let z_{hi} denote the characteristic of interest of the ith observation from the hth stratum and let x_{hi} denote a related concomitant variable. Let $y_{1hi} = x_{hi}$, $y_{2hi} = z_{hi}$, $y_{3hi} = x_{hi}z_{hi}$, $y_{4hi} = x_{hi}^2$, and $y_{5hi} = z_{hi}^2$. The sample estimates of the corresponding population parameters used were: (1) the ratio, $r = \bar y_2/\bar y_1$; (2) the correlation coefficient, $c = [\bar y_3 - \bar y_1\bar y_2/][(\bar y_4 - \bar y_1^2)(\bar y_5 - \bar y_2^2)]^{1/2}$; and (3) the median, $m = P^{-1}(1/2) = \inf\{t : P(t) \ge \frac{1}{2}\}$, where $P(t) = \sum_{h=1}^{L} W_h P_h(t)$ and $P_h(t) = \sum_{i=1}^{n_h} I_{[z_{hi}\le t]}/n_h$, with $\bar y_j = \sum_{h=1}^{L} W_h \bar y_{jh}$ and $\bar y_{jh} = \sum_{i=1}^{n_h} y_{jhi}/n_h$ for $j = 1, \ldots, 5$.

The relative bias and relative stability (defined in the foot-

Table 2. Monte Carlo Rates in the Upper and Lower Tails, Standardized Lengths, Relative Biases, and Relative Stabilities of Six Methods (Median) for Stratified Sampling Without Replacement

	Nominal one-tailed error rate						Standardized length[a]			
	5%			10%						
Method	L	U	L + U	L	U	L + U	5%	10%	% bias[b]	Stab[c]
					$n_h = 5$					
WOOD	6.0	2.2	8.2	5.8	17.6	23.4	1.26	.77		
$\alpha = .01$									220.5	2.46
$\alpha = .025$									−9.2	.44
$\alpha = .05$									18.9	.59
$\alpha = .1$									68.8	1.05
JACK	14.8	6.2	19.0	14.8	13.4	28.2	1.09	1.09	179.3	4.60
PM1	5.8	2.2	8.0	5.8	15.4	21.2	.83	1.04	14.2	.84
PM2	5.8	2.2	8.0	5.8	13.6	19.4	.83	.85	23.2	.95
					$n_h = 7$					
WOOD	0.2	2.6	2.8	6.4	19.0	25.4	1.11	0.69		
$\alpha = .01$									42.7	.65
$\alpha = .025$									−28.5	.46
$\alpha = .05$									−6.4	.47
$\alpha = .1$									32.9	.74
JACK	18.6	8.0	26.6	21.6	8.0	29.6	1.06	1.06	77.6	2.79
PM1	9.0	2.0	11.0	9.0	12.8	21.8	.76	.96	8.5	.81
PM2	1.8	2.2	4.0	9.0	15.8	23.8	1.02	.73	18.2	.79

[a] Standardized length = (length of interval estimate)/($2z_\alpha \sqrt{\text{MSE}(\hat\theta)}$).
[b] % Relative bias = (E(var. est.)/MSE ($\hat\theta$) − 1) × 100%.
[c] Relative stability = $\sqrt{\text{MSE (var. est.)}}$/MSE ($\hat\theta$).

notes to Tables 1 and 2) for the variance estimates and the standardized length (also defined in the footnotes to Tables 1 and 2) and error rates of the confidence intervals (nominal error rates of 5% and 10% in *each* tail) were compared. Bootstrap-t_J intervals (see Section 1.4) were used for the bootstrap methods, and intervals based on asymptotic normality and the variance estimates v_L and v_J were used for the linearization and jackknife methods.

The finite population used for Table 1 consisted of $N = 760$ units, partitioned into $L = 6$ strata, with $N_1 = N_2 = 200$, $N_3 = N_4 = 100$, and $N_5 = N_6 = 80$. It was generated from Kovar, Rao, and Wu's (1988) population 2, which resembled real populations encountered in the National Assessment of Educational Progress study, by first aggregating the hypothetical stratum parameters into six strata and then using a bivariate normal distribution in each stratum with $\rho = .8$, the correlation of x and z, equal across strata. A stratified sample without replacement was taken in each of $S = 1{,}000$ simulations, with $n_h = 20$ for $h = 1, \ldots, 6$. The bootstrap variance estimates and confidence intervals were based on $B = 200$ bootstrap replications. (Sitter [1989] compared using $B = 100$ to $B = 200$ for a subset of the simulations and found no qualitative difference in the results.) Kovar, Rao, and Wu (1988) performed a similar study on linearization, the jackknife, the BRR method, the iid bootstrap, and the rescaling bootstrap for stratified sampling with replacement. They also studied $\theta = g(\bar{\mathbf{y}})$, a nonlinear function of means, and the median.

The results for the ratio and correlation coefficient are presented in Table 1. The mean square error (MSE) $(\hat{\theta})$ was obtained from a separate simulation run of $S = 3{,}000$ samples, with each sample having $n_h = 20$ in each stratum. Some important points to note are as follows:

1. There is little difference in the two-sided $(L + U)$ error rates for the different methods, but the bootstrap methods track one-sided error rates better, more so for the correlation coefficient.
2. The bootstrap methods yield slightly larger standardized lengths, with the rescaling method larger than the proposed method.
3. The methods perform similarly in terms of relative bias of the resulting variance estimates, but the bootstrap methods yield marginally larger relative stability.
4. For both the proposed method and the rescaling method, choosing the resample sizes to match the third moment of $\bar{y}$ yielded marginally lower relative biases.
5. For the proposed method this choice also improved the confidence interval error rates, but there was little difference for the rescaling method in this respect. (This is in contrast with the results of Kovar, Rao, and Wu (1988), where the rescaling bootstrap-t_J performed very poorly when the resample size was chosen to match third moments.)
6. The proposed model (PM1) performed slightly better in terms of coverage probabilities, with the rescaling method having error rates below the nominal levels.

The differences in performance of the methods noted in 1–6, though slight, appeared consistently throughout all the populations of Sitter (1989) with the exception of 4. In general, choosing the resample size to match third moments did not reduce the relative bias for the proposed method. It also should be noted that the poorer relative stability of the bootstrap methods was not as noticeable as it was for sampling with replacement (Kovar, Rao, and Wu 1988). In the larger study of Sitter (1989) some of the finite populations considered were generated using gamma distributions and were more skewed than the one presented in Table 1. The relative performances were not qualitatively different from those in Table 1. The proposed method's (PM1) mirroring of the original sampling scheme seemed to be most beneficial for the median. Choosing n'_h to match third moments did, on average over all the populations, improve the coverage of the confidence intervals for the ratio, regression coefficient, and the correlation coefficient over the simpler BWR (PM2), but perhaps not enough to justify the extra complexity.

The finite population used to generate Table 2 consisted of $N = 800$ units stratified into $L = 32$ strata. The strata sizes ranged from 8 to 256. This population was generated using the hypothetical population based on Hansen, Madow, and Tepping (1983) (stratified by a highly correlated concomitant variable) and used by Kovar, Rao, and Wu (1988). Results for the rescaling method are not presented, because as proposed it is applicable only to θ a smooth function of means.

Table 2 gives results of $S = 500$ simulations with $B = 500$ bootstrap replications for the median using this population. Both $n_h = 5$ and $n_h = 7$ were considered. The MSE of the median was estimated from a separate simulation of $S = 3{,}000$ runs. The results are summarized as follows: Variance estimates based on Woodruff's (1952) method (see Rao and Wu 1987) depend heavily on the choice of α. A poor choice of α can yield very poor variance estimates. The delete-1 jackknife performs poorly, verifying its asymptotic inconsistency for variance estimation of the median. The proposed method (PM1) outperforms the BWR (PM2) and Woodruff method in terms of error rates and relative bias of the variance estimates (unless one is lucky enough to choose a good α value for the Woodruff method). Sitter (1989) considered a number of finite populations for the median. Woodruff's method (with $\alpha = .05$ for variance estimation) and the BWR method performed well, although not quite as well as PM1, for some situations. But they were more sensitive to stratification by a concomitant variable, and in some cases performed much worse than in Table 2.

2. TWO-STAGE CLUSTER SAMPLING WITHOUT REPLACEMENT

In two-stage cluster sampling the population consists of N clusters called primary sampling units (psu), with M_i units in the ith cluster, called secondary sampling units (ssu). A simple random sample of n clusters is drawn without replacement, and then m_i ssu's are drawn without replacement within each obtained cluster. The y_{ij} denotes some measured characteristic, or possibly a vector of measured characteristics, of the jth unit sampled from the ith cluster. Throughout this section we will assume that y_{ij} is univariate for simplicity. The population is denoted by $Y = (Y_1, Y_2, \ldots, Y_N)$, where $Y_i = (y_{i1}, y_{i2}, \ldots, y_{iM_i})$ for $i = 1, 2, \ldots, N$, and the sample is denoted by $y = (y_1, y_2, \ldots, y_n)$, where $y_i = (y_{i1}, y_{i2}, \ldots, y_{im_i})$. The usual unbiased estimator of the population mean is

$$\hat{\bar{Y}} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}\frac{M_i y_{ij}}{\bar{M}_0} = \frac{1}{n}\sum_{i=1}^{n}\frac{M_i\bar{y}_i}{\bar{M}_0}, \qquad (2.1)$$

where $\bar{M}_0 = M_0/N$, $M_0 = \sum_{i=1}^{N} M_i$, and $\bar{y}_i = \sum_{j=1}^{m_i} y_{ij}/m_i$. An unbiased estimate for $\mathrm{Var}(\hat{\bar{Y}})$ was given in Cochran (1977, p. 303) as

$$\mathrm{var}(\hat{\bar{Y}}) = \frac{1-f_1}{n}s_1^2 + \sum_{i=1}^{n}\frac{f_1(1-f_{2i})}{nm_i}s_{2i}^2, \qquad (2.2)$$

where $f_1 = n/N$, $f_{2i} = m_i/M_i$,

$$s_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{M_i\bar{y}_i}{\bar{M}_0} - \hat{\bar{Y}}\right)^2,$$

and

$$s_{2i}^2 = \frac{1}{n(m_i-1)}\sum_{j=1}^{m_i}\left(\frac{M_i}{\bar{M}_0}\right)^2(y_{ij}-\bar{y}_i)^2.$$

2.1 The Proposed Method for Two-Stage Cluster Sampling

The proposed method is extended to two-stage cluster sampling as follows:

1. Choose $1 \le n' < n$ and resample n' clusters from $\{y_i\}_{i=1}^{n}$ without replacement to obtain $\{y_i^*\}_{i=1}^{n'}$, where $y_i^* = (y_{i1}^*, y_{i2}^*, \ldots, y_{im_i}^*)$.
2. Repeat step 1 $k_1 = n(1-f_1^*)/n'(1-f_1)$ times independently to obtain $\{y_i^*\}_{i=1}^{n^*}$, where $n^* = k_1 n'$ and $f_1^* = n'/n$.
3. Choose $1 \le m_i' < m_i$ and resample m_i' units without replacement within the ith cluster obtained in steps 1 and 2 to get $y_i^{**} = (y_{i1}^{**}, y_{i2}^{**}, \ldots, y_{im_i'}^{**})$ for $i = 1, 2, \ldots, n^*$.
4. Repeat step 3 $k_{2i} = [m_i(1-f_{2i}^*)/m_i'(1-f_{2i})]\,(N/n^*)$ times independently, where $f_{2i}^* = m_i'/m_i$, to obtain $y_i^{**} = (y_{i1}^{**}, y_{i2}^{**}, \ldots, y_{im_i^*}^{**})$ for $i = 1, 2, \ldots, n^*$, where $m_i^* = k_{2i}m_i'$.
5. Let $\hat{\theta}^{**} = \hat{\theta}(y_1^{**}, y_2^{**}, \ldots, y_{n^*}^{**})$.
6. Repeat steps 1–5 a large number of times, B, to obtain $\hat{\theta}_1^{**}, \hat{\theta}_2^{**}, \ldots, \hat{\theta}_B^{**}$.
7. Estimate $\mathrm{var}(\hat{\theta})$ with

$$v_{pm} = E_{**}(\hat{\theta}^{**} - E_{**}\hat{\theta}^{**})^2$$

or its Monte Carlo approximation $v_{pm} = \sum_{i=1}^{B}(\hat{\theta}_i^{**} - \hat{\theta}_{(\cdot)}^{**})^2/(B-1)$. In this step, $E_{**}\hat{\theta}^{**}$ and $\hat{\theta}_{(\cdot)}^{**}$ can be replaced by $\hat{\theta}$.

Assume that k_1, and k_{2i} as defined above are integer valued and greater than or equal to 1 (see Section 1.6). In the linear case, $\hat{\theta} = \hat{\bar{Y}}$,

$$v_{pm} = \mathrm{Var}_{**}(\hat{\bar{Y}}^{**}) = \mathrm{var}(\hat{\bar{Y}}),$$

the usual unbiased estimate of $\mathrm{Var}(\hat{\bar{Y}})$. To see this, let E_{2*} and V_{2*} denote the conditional expectation and variance for a given resampled cluster and let E_{1*} and V_{1*} denote the expectation and variance for the resampled clusters. Then

$$\mathrm{Var}_{**}(\hat{\bar{Y}}^{**}) = V_{1*}E_{2*}(\hat{\bar{Y}}^{**}) + E_{1*}V_{2*}(\hat{\bar{Y}}^{**}), \qquad (2.3)$$

with

$$E_{2*}(\hat{\bar{Y}}^{**}) = \frac{1}{n^*}\sum_{i=1}^{n^*}\frac{M_i}{\bar{M}_0}E_{2*}(\bar{y}_i^{**}) = \frac{1}{n^*}\sum_{i=1}^{n^*}\frac{M_i\bar{y}_i}{\bar{M}_0}$$

and

$$V_{2*}(\hat{\bar{Y}}^{**}) = \frac{1}{(n^*)^2}\sum_{i=1}^{n^*}\left(\frac{M_i}{\bar{M}_0}\right)^2 V_{2*}(\bar{y}_i^{**}) = \frac{1}{(n^*)^2}\sum_{i=1}^{n^*}\frac{(1-f_{2i}^*)n}{k_{2i}m_i'}s_{2i}^2 = \frac{1}{n^*}\sum_{i=1}^{n^*}\frac{f_1(1-f_{2i})}{m_i}s_{2i}^2,$$

where

$$s_{2i}^2 = \frac{1}{n(m_i-1)}\sum_{j=1}^{m_i}\left(\frac{M_i}{\bar{M}_0}\right)^2(y_{ij}-\bar{y}_i)^2.$$

So

$$V_{1*}E_{2*}(\hat{\bar{Y}}^{**}) = V_{1*}\left(\frac{1}{n^*}\sum_{i=1}^{n^*}\frac{M_i\bar{y}_i}{\bar{M}_0}\right) = \frac{1-f_1^*}{k_1 n'}s_1^2 = \frac{1-f_1}{n}s_1^2, \qquad (2.4)$$

where

$$s_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{M_i\bar{y}_i}{\bar{M}_0} - \hat{\bar{Y}}\right)^2,$$

and

$$E_{1*}V_{2*}(\hat{\bar{Y}}^{**}) = E_{1*}\left(\frac{1}{n^*}\sum_{i=1}^{n^*}\frac{f_1(1-f_{2i})}{m_i}s_{2i}^2\right) = \sum_{i=1}^{n}\frac{f_1(1-f_{2i})}{nm_i}s_{2i}^2. \qquad (2.5)$$

Substituting (2.4) and (2.5) into (2.3) yields the result. If the integer assumptions in the algorithm do not hold, a randomization between bracketing integers similar to that of Section 1.6 is available (Sitter 1989).

Rao and Wu (1988) gave an extension of the rescaling method to two-stage sampling that involves rescaling at each stage. It is difficult to say whether the method proposed in this article has any advantages over the rescaling method for two-stage cluster sampling. Both are complex algorithms in which performance is difficult to compare, either theoretically or via simulation. Choosing resample size to match Edgeworth expansions is unavailable, because Edgeworth expansions for two-stage sampling have not yet been developed. McCarthy and Snowden (1985) discussed extensions of the BWR and BWO methods to two-stage sampling in the special case of equal cluster sizes, $M_i = M$, and equal within-cluster sample sizes, $m_i = m$. The BWR extension is discussed as a special case of the proposed method in Section 2.2. The BWO extension derived by McCarthy and Snowden is shown to not match the usual variance estimate in the linear case except under some very special situations. Sitter (1992) derived a BWO extension that does not suffer this limitation.

2.2 Extension of the BWR Method to Two-Stage Cluster Sampling With Unequal Cluster Sizes

In stratified random sampling the proposed method allows any choice of $1 \le n'_h < n_h$ for each h (see Section 1.2). Suppose one chooses $n'_h = 1$ for all h; then the algorithm becomes a resampling method with replacement, and total resample size is $(n_h - 1)/(1 - f_h)$. This resample size is the same as that of the BWR method. So choosing $n'_h = 1$ for each h reduces the proposed method to the BWR method for stratified sampling. If the same idea is now applied to the proposed method in the context of two-stage cluster sampling, a BWR method for two-stage cluster sampling can be obtained. In the proposed algorithm, choose $n' = 1$ and $m'_i = 1$ for each i. The resulting algorithm can be written as follows:

1. Resample $n^* = (n-1)/(1-f_1)$ clusters with replacement.
2. Within each cluster obtained in step 1, resample $m_i^* = (m_i - 1)N/(1 - f_{2i})n^*$ units with replacement to get $y_i^{**} = (y_1^{**}, y_2^{**}, \ldots, y_{m_i^*}^{**})$ for $i = 1, 2, \ldots, n^*$.
3. Let $\hat{\theta}^{**} = \hat{\theta}(y_1^{**}, \ldots, y_{n^*}^{**})$.
4. Repeat steps 1–3 a large number of times, B, to obtain $\hat{\theta}_1^{**}, \ldots, \hat{\theta}_B^{**}$.
5. Estimate $\text{var}(\hat{\theta})$ with

$$v_{bwr} = E_{**}(\hat{\theta}^{**} - E_{**}\hat{\theta}^{**})^2$$

or its Monte Carlo approximation $v_{pm} = \sum_{i=1}^{B} (\hat{\theta}_i^{**} - \hat{\theta}_{(\cdot)}^{**})^2/(B-1)$. In this step, $E_{**}\hat{\theta}^{**}$ and $\hat{\theta}_{(\cdot)}^{**}$ can be replaced by $\hat{\theta}$.

Assume that n^* and m_i^* are integer valued (see Section 1.6). Because this indeed is a special case of the proposed method, with $\hat{\theta} = \hat{\bar{Y}}$, $v_{bwr} = \text{Var}_{**}(\hat{\bar{Y}}^{**}) = \text{var}(\hat{\bar{Y}})$. This is the logical extension of the BWR method for two-stage sampling with equal cluster sizes proposed by McCarthy and Snowden (1985) to the unequal cluster size situation, and in the case where $M_i = M$, and $m_i = m$ it reduces to their method.

3. THE RAO–HARTLEY–COCHRAN METHOD (RHC) FOR PPS SAMPLING

A simple method of sampling with unequal probabilities without replacement was proposed by Rao, Hartley, and Cochran (1962); see also Cochran (1977). The sampling method is as follows:

1. Randomly partition the population of N units into n groups $\{G_g\}_{g=1}^{n}$ of sizes $\{N_g\}_{g=1}^{n}$.
2. Draw one unit from each group with probability z_j/Z_g for the gth group, where $z_j = x_j/X$, $Z_g = \sum_{j \in G_g} z_j$, x_j = some measure of the size of the jth unit, and $X = \sum_{j=1}^{N} x_j$.

An unbiased estimator of $\bar{Y}$, the population mean, is

$$\hat{\bar{Y}} = \frac{1}{n} \sum_{g=1}^{n} w_g y_g,$$

where $w_g = f/\pi_g$, $\pi_g = z_g/Z_g$ is the probability of selection of the gth sampled unit, $f = n/N$ is the probability of selection under simple random sampling without replacement, and $y_g = z_g$ denote the values for the unit selected from the gth group. Note that by definition, $\sum_{g=1}^{n} Z_g = 1$. An unbiased estimate of $\text{Var}(\hat{\bar{Y}})$ is

$$\text{var}(\hat{\bar{Y}}) = \frac{(\sum_{g=1}^{n} N_g^2 - N)}{(N^2 - \sum_{g=1}^{n} N_g^2)} \sum_{g=1}^{n} Z_g \left(\frac{y_g}{N z_g} - \hat{\bar{Y}} \right)^2.$$

The proposed method is extended to this situation in the following manner:

1. Let $v_i = w_i y_i$ for $i = 1, 2, \ldots, n$.
2. Choose $1 \le n^* < n$, an integer, and randomly partition the v_i's into n^* groups, $\{\Gamma_g^*\}_{g=1}^{n^*}$, of sizes $\{n_g\}_{g=1}^{n^*}$.
3. Randomly select one v_i from each of the n^* groups with probability z_g/z_{tg}, where $Z_{tg} = \sum_{i \in \Gamma_g^*} Z_i$, to get $v_1^*, v_2^*, \ldots, v_{n^*}^*$.
4. Repeat steps 2–3

$$k = \frac{(\sum_{g=1}^{n} n_g^2 - n)}{n(n-1)} \cdot \frac{(N^2 - \sum_{g=1}^{n} N_g^2)}{(\sum_{g=1}^{n} N_g^2 - N)}$$

times, independently, replacing the v_i's drawn at each step. If k is noninteger valued, a randomization between bracketing integers is available. Let $\hat{\theta}^* = \hat{\theta}(v_1^*, v_2^*, \ldots, v_{kn^*}^*)$.

5. Repeat steps 2–4 a large number of times, B, to get $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$.
6. Estimate $\text{Var}(\hat{\theta})$ with

$$v_{pm} = E_*(\hat{\theta}^* - E_*\hat{\theta}^*)^2$$

or its Monte Carlo approximation $v_{pm} = \sum_{i=1}^{B} (\hat{\theta}_i^* - \hat{\theta}_{(\cdot)}^*)^2/B$. In this step, $E_*\hat{\theta}^*$ and $\hat{\theta}_{(\cdot)}^*$ can be replaced by $\hat{\theta}$.

For any choices of n^* and n_g's that allow this algorithm, if $\theta = \bar{Y}$ and $\hat{\theta} = \hat{\bar{Y}}$ then

$$v_{pm} = \text{var}(\hat{\bar{Y}}).$$

To show this, note that

$$\hat{\bar{Y}}^* = \frac{1}{kn^*} \sum_{i=1}^{k} \sum_{g=1}^{n^*} \frac{f^* Z_{tgi}}{z_{gi}} v_{gi}$$

and

$$E_*(\hat{\bar{Y}}^*) = \frac{1}{n} \sum_{g=1}^{n} v_g = \hat{\bar{Y}},$$

where the subscript i represents the with-replacement step of the algorithm. Thus $v_{pm} = \text{Var}_*(\hat{\bar{Y}}^*)$ and, using the well-known result for the Rao–Hartley–Cochran method (see Cochran 1977),

$$\operatorname{Var}_*(\hat{\hat{Y}}^*) = \frac{1}{kn^2}\frac{(\sum_{g=1}^{n^*} n_g^2 - n)}{n(n-1)}\left(\sum_{g=1}^{n^*}\frac{v_g^2}{Z_g} - \left(\sum_{g=1}^{n^*} v_g\right)^2\right)$$
$$= \frac{1}{k}\frac{(\sum_{g=1}^{n^*} n_g^2 - n)}{n(n-1)}\sum_{g=1}^{n^*} Z_g\left(\frac{y_g}{Nz_g} - \hat{\hat{Y}}\right)^2.$$

Substituting for k yields the result.

If in the original sampling, $N_g = N/n$ and $z_g = n/N$ for $g = 1, \ldots, n$, then the Rao–Hartley–Cochran scheme reduces to simple random sampling without replacement. If in this case the n_g's are chosen equal, then the method reduces to the proposed method for simple random sampling without replacement.

4. CONCLUSIONS

This article develops extensions of the concept of resampling to sample survey data. The proposed resampling methods are applicable to stratified multistage sample designs, as well as to the Rao–Hartley–Cochran method of unequal probability sampling. They yield variance estimators for a general class of estimators, including nonlinear functions of means and percentiles. All of the variance estimates reduce to the standard ones in the linear case. The methods also allow the estimation of confidence intervals for $\hat{\theta}$ that take into account the skewness of its distribution. There is theoretical evidence of this in terms of Edgeworth expansions, as well as some empirical evidence in that the confidence interval estimates track the one-tailed error rates better than do the jackknife and linearization methods. There also is empirical evidence that the proposed methods perform better for percentile estimation in stratified samples, both for variance estimation and for confidence intervals.

APPENDIX A: PROOF OF (1.3)

Assume

$$\max_h \frac{W_h}{n_h} = O(n^{-1}), \qquad 0 \le f_h, \qquad f_h^* \le \delta < 1$$
$$h = 1, \ldots, L. \quad \text{(A.1)}$$

Let $\bar{\mathbf{z}}^* = \bar{\mathbf{y}}^* - \bar{\mathbf{y}} = (\bar{y}_1^* - \bar{y}_1, \ldots, \bar{y}_\tau^* - \bar{y}_\tau)^T$. The assumption (A.1), together with the boundedness of $S_{h\alpha}^2 = E(y_{h\alpha} - \bar{Y}_{h\alpha})^2$, imply

$$\operatorname{var}_*(\bar{z}_\alpha^*) = \sum_{h=1}^{L} W_h \frac{(1-f_h)}{n_h} s_{\alpha h}^2 = O_p(n^{-1})$$

for $\alpha = 1, \ldots, \tau$. So the elements of $\bar{\mathbf{z}}^*$ are of order $n^{-1/2}$ in probability $[O_p(n^{-1/2})]$. Let $g'(\bar{\mathbf{y}}) = [g_1'(\bar{\mathbf{y}}), \ldots, g_\tau'(\bar{\mathbf{y}})]^T$ be the vector of first derivatives $\partial g(\mathbf{x})/\partial x_\alpha$ evaluated at $\mathbf{x} = \bar{\mathbf{y}}$. Let $g''(\bar{\mathbf{y}}) = \{g_{\alpha\beta}(\bar{\mathbf{y}})\}$ be the $\tau \times \tau$ matrix of second derivatives $\partial g^2(\mathbf{x})/\partial x_\alpha \partial x_\beta$ evaluated at $\mathbf{x} = \bar{\mathbf{y}}$. Then

$$\hat{\theta}^* = g(\bar{\mathbf{y}}^*) = \hat{\theta} + \bar{\mathbf{z}}^{*T} g'(\bar{\mathbf{y}}) + \frac{1}{2}\bar{z}^{*T} g''(\bar{\mathbf{y}})\bar{\mathbf{z}}^* + O_p(n^{-3/2}). \quad \text{(A.2)}$$

Therefore,

$$E_*(\hat{\theta}^* - \hat{\theta})^2 = \sum_{\alpha,\beta=1}^{\tau} \alpha_\alpha(\bar{\mathbf{y}}) g_\beta(\bar{\mathbf{y}}) E_*(\bar{z}_\alpha^* \bar{z}_\beta^*)$$
$$+ \sum_{\alpha,\beta,\gamma=1}^{\tau} g_\alpha(\bar{\mathbf{y}}) g_{\beta\gamma}(\bar{\mathbf{y}}) E_*(\bar{z}_\alpha^* \bar{z}_\beta^* \bar{z}_\gamma^*) + O_p(n^{-2}).$$

This step is not strictly rigorous. To make it so, one must impose some regularity conditions on g and higher moments of $\bar{\mathbf{y}}$ and apply Theorem 5.4.3 of Fuller (1976). Now note that

$$E_*(\bar{z}_\alpha^* \bar{z}_\beta^*) = \sum_{h=1}^{L} W_h^2 \frac{(1-f_h)}{n_h} s_{\alpha\beta h}, \quad \text{(A.3)}$$

where $(n_h - 1)s_{\alpha\beta h} = \sum_i (y_{\alpha hi} - \bar{y}_{\alpha h})(y_{\beta hi} - \bar{y}_{\beta h})$, and

$$E_*(\bar{z}_\alpha^* \bar{z}_\beta^* \bar{z}_\gamma^*) = \sum_{h=1}^{L} W_h^3 \frac{(1-f_h)^2(1-2f_h^*)}{(1-f_h^*)(n_h-1)(n_h-2)} s_{\alpha\beta\gamma h}$$
$$= O_p(n^{-2}),$$

where $(n_h - 1)s_{\alpha\beta\gamma h} = \sum_i (y_{\alpha hi} - \bar{y}_{\alpha h})(y_{\beta hi} - \bar{y}_{\beta h})(y_{\gamma hi} - \bar{y}_{\gamma h})$. Thus the result follows.

APPENDIX B: PROOF OF (1.4)

Let $\varepsilon_h = W_h(\bar{y}_h^* - \bar{y}_h)$ for $h = 1, \ldots, L$. Thus $E_*(\varepsilon_h) = 0$, and then

$$E_*(\bar{y}^* - \bar{y})^3 = E_*\left(\sum_{h=1}^{L} \varepsilon_h\right)^3$$
$$= E_*\left(\sum_{h=1}^{L} \varepsilon_h^3 + 3\sum_{h\ne l}^{L} \varepsilon_h \varepsilon_l^2 + \sum_{h\ne l\ne j}^{L} \varepsilon_h \varepsilon_l \varepsilon_j\right)$$
$$= \sum_{h=1}^{L} E_*(\varepsilon_h^3) = \sum_{h=1}^{L} W_h^3 E_*(\bar{y}_h^* - \bar{y}_h)^3. \quad \text{(B.1)}$$

We need find only $E_*(\bar{y}_h^* - \bar{y}_h)^3$. Suppressing the h subscript, we can write $\bar{y}^* = \sum_{j=1}^{k} \bar{y}_j^*/k = \sum_{j=1}^{k}\sum_{l=1}^{n'} y_{jl}^*/kn'$, where the j and l subscripts refer to the with-replacement step and the without-replacement step. So

$$E_*(\bar{y}^* - \bar{y})^3$$
$$= \frac{1}{k^3}\left(\sum_{j=1}^{k} E_*(\bar{y}_j^* - \bar{y})^3 + 3\sum_{j\ne l}^{k} E_*(\bar{y}_j^* - \bar{y})E_*(\bar{y}_l^* - \bar{y})^2 + \sum_{j\ne l\ne m}^{k} E_*(\bar{y}_j^* - \bar{y})E_*(\bar{y}_l^* - \bar{y})E_*(\bar{y}_m^* - \bar{y})\right)$$
$$= \frac{1}{k}\sum_{j=1}^{k} E_*(\bar{y}_j^* - \bar{y})^3$$
$$= \frac{n^2(1-f^*)(1-2f^*)m_3}{(kn')^2(n-1)(n-2)}, \quad \text{(B.2)}$$

where the last step follows from the fact that $\bar{y}_j^*$ was obtained by SRSWOR of n' units from the original sample of size n and using the results of Sukhatme and Sukhatme (1970) on higher moments of the mean from a SRSWOR. All that remains is to note that

$$\frac{n_h^2(1-f_h^*)(1-2f_h^*)m_{3h}}{(k_h n_h')^2(n_h-1)(n_h-2)} = \frac{(1-f_h)(1-2f_h)m_{3h}}{(n_h-1)(n_h-2)}[1 - A_h],$$

where A_h is given in (1.5).

[*Received June 1990. Revised August 1991.*]

REFERENCES

Abromovitch, L., and Singh, K. (1985), "Edgeworth Corrected Pivotal Statistics and the Bootstrap," *The Annals of Statistics,* 13, 116–132.

Babu, G. J., and Singh, K. (1985), "Edgeworth Expansions for Sampling Without Replacement From Finite Populations," *Jounral of Multivariate Analysis,* 17, 261–278.

Bickel, P. J., and Freedman, D. A. (1984), "Asymptotic Normality and the Bootstrap in Stratified Sampling," *The Annals of Statistics,* 12, 470–482.

Chen, J., and Sitter, R. R. (1991), "Edgeworth Expansions and the Bootstrap for Stratified Sampling Without Replacement From a Finite Population," Technical Report 163, Carleton University, Laboratory for Research in Statistics and Probability.

Cochran, W. G. (1977), *Sampling Techniques,* New York: John Wiley.

Feller, W. (1966), *An Introduction to Probability Theory and Its Applications: Volume II* (2nd ed.), New York: John Wiley.

Fuller, W. A. (1976), *Introduction to Statistical Time Series,* New York: John Wiley.

Gross, S. (1980), "Median Estimation in Sample Surveys," in *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 181–184.

Hansen, M. H., Madow, W. G., and Tepping, B. J. (1983), "An Evaluation of Model-Dependent and Probability-Sampling Inferences in Sample Survey," *Journal of the American Statistical Association,* 78, 776–807.

Kovar, J. G., Rao, J. N. K., and Wu, C. F. J. (1988), "Bootstrap and Other Methods to Measure Errors in Survey Estimates," *The Canadian Journal of Statistics,* 16, Supplement, 25–45.

Krewski, D., and Rao, J. N. K. (1981), "Inference From Stratified Samples: Properties of the Linearization, Jackknife, and Balanced Repeated Replication Methods," *The Annals of Statistics,* 9, 1010–1019.

McCarthy, P. J., and Snowden, C. B. (1985), "The Bootstrap and Finite Population Sampling," in *Vital and Health Statistics* (Ser. 2, No. 95), Public Health Service Publication, Washington, DC: U.S. Government Printing Office, pp. 85–1369.

Rao, J. N. K., Hartley, H. O., and Cochran, W. G. (1962), "On a Simple Procedure of Unequal Probability Sampling Without Replacement," *Journal of the Royal Statistical Society,* Ser. B, 24, 482–491.

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.

——— (1987), "Methods for Standard Errors and Confidence Intervals From Sample Survey Data: Some Recent Work," in *The Bulletin of the International Statistical Institute,* Proceedings of the 46th Session, Book 3, pp. 5–19.

——— (1988), "Resampling Interference With Complex Survey Data," *Journal of the American Statistical Association,* 83, 231–241.

Robinson, J. (1978), "An Asymptotic Expansion for Samples From a Finite Population," *The Annals of Statistics,* 6, 1005–1011.

Sitter, R. R. (1989), "Resampling Procedures for Complex Survey Data," unpublished Ph.D. dissertation, University of Waterloo.

——— (1992), "Comparing Three Bootstrap Methods for Survey Data," *The Canadian Jounal of Statistics,* 20, No. 2, 135–154.

Sukhatme, P. V., and Sukhatme, B. W. (1970), *Sampling Theory of Surveys With Applications,* Ames IA: Iowa State University Press.

Woodruff, R. S. (1952), "Confidence Intervals for Medians and Other Position Measures," *Journal of the American Statistical Association,* 47, 635–646.