SOME PREDICTION PROPERTIES OF BALANCED HALF-SAMPLE VARIANCE ESTIMATORS IN SINGLE-STAGE SAMPLING

Richard Valliant, U.S. Bureau of Labor Statistics

1. INTRODUCTION

Estimation of variances in complex surveys is often done by some type of replication. One of the most common is the balanced half-sample (BHS) or balanced repeated replication method originated by McCarthy (1966, 1969a, 1969b). A number of United States government surveys have used BHS including the Producer Price Index, the Consumer Expenditure Survey, and the Employment Cost Index, all sponsored by the U.S. Bureau of Labor Statistics, the Survey of Income and Program Participation and the 1980 Census Post Enumeration Survey conducted by the U.S. Bureau of the Census, and the National Health Interview Survey and the Health and Nutrition Examination Survey sponsored by the U.S. National Center for Health Statistics.

Recently, a number of theoretical studies of BHS have been conducted from the point of view of probability sampling theory including Krewski (1978), Krewski and Rao (1981), Dippo (1981), Dippo and Wolter (1984), and Rao and Wu (1985). Prior to these works, a variety of empirical studies were done and are listed in Krewski and Rao. These theoretical and empirical studies have concentrated on randomization properties such as asymptotic design-unbiasedness and design-consistency.

Often there is a well-defined reference set which may used in making conditional inferences once the data have been observed (Royall 1976, Smith 1984). The prediction theoretic approach to finite population inference has proved to be a valuable means of conditioning on the achieved sample rather than averaging over all possible samples. In some common situations, conditional and unconditional properties of estimators of totals and associated variance estimators can be dramatically different (Royall and Cumberland 1978, 1981a, 1981b, 1985; Holt and Smith 1979; Cumberland and Royall 1981). Consequently, an important consideration is whether the BHS variance estimator has good conditional properties in addition to the desirable unconditional properties demonstrated by the earlier studies.

Implementation of the BHS method many times involves collapsing of strata or grouping of primary sampling units (psu's) in order to form pairs. A set of half-samples from the paired groups is then selected. The collapsing is generally done to (1) save computation time by reducing the number of half-samples and (2) adapt the BHS method to surveys which do not use two-psus-per-stratum designs. The pairs may be formed by combining two strata in a one-psu-per-stratum design or by grouping psu's within a stratum when more than two psu's are sampled. In this paper we consider the latter approach which was also studied by Krewski (1978) and Dippo (1981). This approach contrasts to that of Krewski and Rao (1981) and Rao and Wu (1985) who derived asymptotic properties of BHS variance estimators when two psu's were selected per stratum and the number of strata was large. The approach using grouping of psu's allows approximate theory to be developed for certain estimators, such as the separate ratio and regression estimators, which are not in the class covered by the last two studies mentioned above and is also similar to methods often used in practice.

Section 2 introduces a prediction model and a class of estimators which includes several ratio, regression, and other estimators which are often studied in finite population inference. Section 3 gives an approximation to the BHS estimator and compares it to the linearization and jackknife estimators. Prediction theory for the BHS estimator as applied to the separate ratio and regression estimators is sketched in section 4. Section 5 reports the results of a simulation study in which the performance of the BHS estimator is compared to that of the linearization and jackknife estimators both in terms of mean squared error estimation and confidence interval construction. We conclude with a brief summary.

2. THE PREDICTION MODEL AND TWO ESTIMATORS OF TOTALS

This section introduces a prediction model and describes a general class of estimators of totals and two specific cases which will be studied. The situation considered here is one in which the units in the population are organized into H strata. A sample of n_h units is selected from the total of N_h units in stratum h. The total number of units in the population is denoted by $N=\Sigma_h^H N_h$ and the total in the sample by $n=\Sigma_h^H n_h$. The set of sample units in stratum h is denoted by s_h.

Associated with each unit (hi) in the finite population is a random variable y_{hi} whose population total is $T = \Sigma_h^H \Sigma_i^{N_h} y_{hi}$ and an auxiliary variable x_{hi} whose value is known for all units in the population. The general prediction model we will consider is :

$$E_\xi(y_{hi}) = \alpha_h + \beta_h x_{hi} \qquad (1)$$
$$\mathrm{var}_\xi(y_{hi}) = v_{hi}$$

with the y_{hi} uncorrelated. The type of estimator that will be considered has the general form

$$\hat{T} = \Sigma_h \Sigma_{i\in s_h} \gamma_{hi} y_{hi} \qquad (2)$$

where γ_{hi} is a constant with respect to the prediction model. A further condition that we require of $\hat{T}$ is that it be a function of within-stratum means. In particular, define the average per sample unit $\bar{z}_{kh} = \Sigma_{i\in s_h} z_{khi}/n_h$ where z_{khi} is a quantity (random or fixed with respect to the model) associated with sample unit (hi) and k = 1,2,...,K. The H-vector of the averages for quantity k is $\bar{\underset{\sim}{z}}_k = (\bar{z}_{k1},\bar{z}_{k2},\ldots,\bar{z}_{kH})'$. The estimator $\hat{T}$ is required to be a function of $\bar{\underset{\sim}{z}}_1$, $\bar{\underset{\sim}{z}}_2,\ldots, \bar{\underset{\sim}{z}}_K$, i.e. $\hat{T} = \hat{T}(\bar{\underset{\sim}{z}}_1,\bar{\underset{\sim}{z}}_2,\ldots,\bar{\underset{\sim}{z}}_K)$. To clarify the coverage of the class of estimators which are given by (2) and are also functions of $\bar{z}_k$, we give two examples.

Among the estimators in this class are two which are normally studied under a stratified simple random sampling (STSRS) plan and are ordinarily thought of as nonlinear in probability sampling theory. They are the separate ratio and linear regression estimators defined as

$$\hat{T}_R = \Sigma_h \ (N_h \bar{y}_{hs} \bar{x}_h / \bar{x}_{hs}) \text{ and}$$

$$\hat{T}_{LR} = \Sigma_h \ N_h[\bar{y}_{hs} + b_{hs}(\bar{x}_h - \bar{x}_{hs})] \text{ , where}$$

$\bar{y}_{hs} = \Sigma_{i\in s_h} y_{hi}/n_h$, $\bar{x}_{hs} = \Sigma_{i\in s_h} x_{hi}/n_h$, $\bar{x}_h = \Sigma_i^{N_h} x_{hi}/N_h$, and $b_{hs} = \Sigma_{i\in s_h}(x_{hi}-\bar{x}_{hs})y_{hi}/\Sigma_{i\in s_h}(x_{hi}-\bar{x}_{hs})^2$. The estimator $\hat{T}_R$, for example, is seen to be in the class of estimators described above by defining $\gamma_{hi} = N_h\bar{x}_h/(n_h\bar{x}_{hs})$, $\bar{z}_{1h} = \bar{y}_{hs}$, and $\bar{z}_{2h} = \bar{x}_{hs}$. The stratified expansion estimator, the combined ratio estimator, and the combined linear regression estimator are also in the class but will not be considered further here.

Model (1) will often be a reasonable approximation to a smooth regression relationship between y and a measure of size x when units are stratified based on x. Under (1) $\hat{T}_{LR}$ is ξ-unbiased while $\hat{T}_R$ is if $\alpha_h = 0$, where h=1,2,...,H. $\hat{T}_R$ will also be ξ-unbiased if a stratified sample balanced on x is selected, i.e. if $\bar{x}_{hs} = \bar{x}_h$ in all strata. Although the fact that $\hat{T}_R$ is generally ξ-biased under (1) argues against its indiscriminate use, it is still of some interest to study the properties of the BHS method for $\hat{T}_R$ because it is widely employed in practice.

The prediction-variance of the general $\hat{T}$ is defined as $var_\xi(\hat{T}-T)$ where the variance is taken with respect to model (1). If as N_h and $n_h \to \infty$, N_h/N and n_h/n converge to constants and other mild regularity conditions hold on various population and sample quantities, the prediction-variance is $O(N^2/n)$ and is asymptotically equivalent to $var_\xi(\hat{T})$:

$$var_\xi(\hat{T}-T) \approx var_\xi(\hat{T}) = \Sigma_h\Sigma_{i\in sh} \ \gamma_{hi}^2 v_{hi} \ . \qquad (3)$$

3. THE BALANCED HALF-SAMPLE VARIANCE ESTIMATOR AND TWO ALTERNATIVES

One measure of the worth of the BHS estimator with grouping of units is its ability to estimate the asymptotic prediction-variance given by (3). To implement the BHS estimator, suppose that the sample units in each stratum are divided into two groups each of which contains $n_h/2$ units where for simplicity n_h is taken to be even. How the groups are formed is an important consideration and will be addressed in later sections. A set of J half-samples may be defined by the indicators

$$\delta_{hg\alpha} = \begin{cases} 1 \text{ if group g in stratum h is in half-sample } \alpha, \\ 0 \text{ if not} \end{cases}$$

for g = 1,2 and α = 1,2,...,J. Based on the $\delta_{hg\alpha}$ define

$$\delta_h^{(\alpha)} = 2\delta_{h1\alpha} - 1 = \begin{cases} 1 \text{ if group 1 in stratum h is in half-sample } \alpha \\ -1 \text{ if group 2 in stratum h is in half-sample } \alpha \end{cases}$$

and note that $-\delta_h^{(\alpha)} = 2\delta_{h2a} - 1$. A set of half-samples is orthogonally balanced on the groups if $\Sigma_\alpha^J \delta_h^{(\alpha)} = \Sigma_\alpha^J \delta_h^{(\alpha)}\delta_{h'}^{(\alpha)} = 0$, $h \neq h'$. A minimal set of balanced half-samples has $H < J \leq H + 4$. Define $t_{khg} = \Sigma_{i\in G_{hg}} z_{khi}$, the total of z_{khi} for the set, denoted G_{hg}, of sample units in group g (g = 1,2) within stratum h. Further, we define $\bar{z}_{kh}^{(\alpha)} = 2(\delta_{h1\alpha}t_{kh1} + \delta_{h2\alpha}t_{kh2})/n_h$ which is the analog to $\bar{z}_{kh}$ based on half-sample α, and let $\hat{T}^{(\alpha)} = \hat{T}(\bar{z}_1^{(\alpha)},\ldots,\bar{z}_K^{(\alpha)})$ where $\bar{z}_k^{(\alpha)} = (\bar{z}_{k1}^{(\alpha)},\ldots,\bar{z}_{kH}^{(\alpha)})'$. The BHS variance estimator to be studied here is

$$v_B(\hat{T}) = \Sigma_\alpha^J \ [\hat{T}^{(\alpha)} - \hat{T}]^2/J. \qquad (4)$$

Alternative BHS variance estimators employing the complement to each half-sample have been suggested by McCarthy and others, but in practice $v_B(\hat{T})$ is used most often.

When n_h is large in all strata and units are allocated to the two groups in such a way that $\hat{T}^{(\alpha)}$ may be reasonably approximated by a first-order Taylor series expanded about $(\bar{z}_1,\ldots,\bar{z}_K)$, then v_B can be approximated as

$$v_B(\hat{T}) \doteq \Sigma_h(\Sigma_k \ d_{kh}\lambda_{kh})^2/n_h^2 \qquad (5)$$

where $d_{kh}=t_{kh1}-t_{kh2}$ and $\lambda_{kh}=\partial\hat{T}^{(\alpha)}/\partial\bar{z}_{kh}^{(\alpha)}$ evaluated at $\bar{z}_{kh}^{(\alpha)}=\bar{z}_{kh}$ for all k and h. Further details are given in Valliant (1987a). The approximate ξ-expectation of $v_B(\hat{T})$ will be evaluated for specific estimators in section 4.

Two competitors to the BHS method are the linearization estimator, denoted here by v_L, and the jackknife estimator, denoted by v_J. The linearization estimator, in the notation of this paper, is

$$v_L = \Sigma_h(1 - f_h)\Sigma_{k,k'}\lambda_{kh}\lambda_{k'h}S_{hkk'}/n_h \quad (6)$$

where $S_{hkk'} = \Sigma_{i\in s_h}(z_{khi}-\bar{z}_{kh})(z_{k'hi}-\bar{z}_{k'h})/(n_h-1)$. Usually, v_L is considered as an estimator of the probability sampling variance, $var_p(\hat{T}-T)$ where the subscript p denotes the distribution induced by the sampling plan. The stratified jackknife estimator as defined by Jones (1974) is

$$v_J = \Sigma_h(1-f_h)(n_h-1)n_h^{-1}\Sigma_{s_h}[\hat{T}_{(hi)}-\hat{T}_{(h)}]^2 \quad (7)$$

where $\hat{T}_{(hi)}$ has the same form as $\hat{T}$ but omits sample unit (hi) and $\hat{T}_{(h)} = \Sigma_{s_h}\hat{T}_{(hi)}/n_h$.

In a different asymptotic framework in which $n_h = 2$ and $H \to \infty$, Rao and Wu (1985) have shown that v_B, v_L, and v_J are equivalent to a first-order approximation. A similar situation occurs here when units are randomly allocated to the two groups within each stratum for v_B. The results are sketched below.

Assignment of units to groups by simple random sampling (SRS) is the conventional method when the goal is to estimate the $var_p(\hat{T}-T)$ (Kish and Frankel 1970, Krewski 1978, Rust 1985). Let r denote the distribution induced by forming half-samples by SRS. It follows that $E_r(d_{kh}) = 0$ and $cov_r(d_{kh},d_{k'h}) = n_hS_{hkk'}$. Consequently, $E_r(v_B) = v_L$, assuming $f_h = 0$ in (6), and, generally, $v_B = v_L + O_r(N^2/n^{1.5})$. For the jackknife use the approximation $\hat{T}_{(hi)} \doteq \hat{T} + \Sigma_k\lambda_{kh}(\bar{z}_{kh(i)}-\bar{z}_{kh})$ where $\bar{z}_{kh(i)}$ is the stratum mean omitting unit (hi). Since $\bar{z}_{kh(i)}-\bar{z}_{kh} = (\bar{z}_{kh}-z_{khi})/(n_h-1)$ we have

$$v_J \doteq \Sigma_h\Sigma_{s_h}[\Sigma_k\lambda_{kh}(z_{khi} - \bar{z}_{kh})]^2/[n_h(n_h - 1)] = v_L ,$$

assuming $f_h = 0$ in (6) and (7). Thus, unconditional on the r-distribution v_B, v_L, and v_J should perform similarly in large samples. However, in small or moderate size samples there may be important differences among the three as noted in sections 4 and 5.

4. THEORY FOR PARTICULAR ESTIMATORS OF TOTALS

This section addresses approximate prediction properties of the estimators used in single-stage sampling described in section 2. Prediction properties are ones computed based on the model (ξ-distribution) and are conditional on the particular sample selected and the particular assignment of units to groups within strata, i.e. conditional on both the p and r-distributions. These conditional properties often contrast to those obtained by averaging over the p, ξ, and r distributions. For cases in which an estimator $\hat{T}$ is ξ-biased under (1), none of the variance estimators considered here does a reasonable job of estimating the prediction mean squared error (MSE), $E_\xi(\hat{T}-T)^2$. However, estimation of the average (p,ξ)-MSE, $E_pE_\xi(\hat{T}-T)^2$, is often possible. In fact, in each of the cases considered here the p-distribution is STSRS and $E_p(v_L) \doteq E_p(\hat{T}-T)^2$ implying that $E_pE_\xi(v_L) \doteq E_pE_\xi(\hat{T}-T)^2$, and a similar approximation holds for v_J. Because $E_r(v_B) \doteq v_L$ when SRS group assignment is used for v_B, we also obtain $E_pE_\xi E_r(v_B) \doteq E_pE_\xi(\hat{T}-T)^2$. Thus, unconditionally v_B, v_L, and v_J are approximately unbiased estimators of the (p,ξ)-MSE.

The prediction properties under model (1) of the linearization and jackknife estimators are given in Royall and Cumberland (1978, 1981a,b) and Valliant (1987b) and are briefly summarized here.

In the special case of (1) in which a particular $\hat{T}$ is ξ-unbiased the jackknife is robust in the sense of being asymptotically ξ-unbiased for the prediction-variance regardless of the specification of v_{hi} in the model. If $\alpha_h = 0$ in (1), then $\hat{T}_R$ is ξ-unbiased but the approximate ξ-bias of v_L is dependent on $1 - (\bar{x}_h/\bar{x}_{hs})^2$. Under (1) $\hat{T}_{LR}$ is ξ-unbiased and v_L has a large-sample ξ-bias dependent on $\bar{x}_{hs} - \bar{x}_h$ although the ξ-bias is a lower order of magnitude than the prediction-variance of $\hat{T}_{LR}$ if an STSRS plan is used.

4.1 Separate Ratio Estimator

The approximate ξ-variance of $\hat{T}_R$ is $\Sigma_h(N_h^2/n_h)(\bar{x}_h/\bar{x}_{hs})^2\bar{v}_{hs}$ where $\bar{v}_{hs} = \Sigma_{s_h}v_{hi}/n_h$. Next, define $d_{1h} = n_h[\bar{y}_{hs1}-\bar{y}_{hs2}]/2$, $d_{2h} = n_h[\bar{x}_{hs1}-\bar{x}_{hs2}]/2$, $\bar{z}_{1h} = \bar{y}_{hs}$, and $\bar{z}_{2h} = \bar{x}_{hs}$ where $\bar{y}_{hsg} = 2\Sigma_{i\in G_{hg}}y_{hi}/n_h$ and $\bar{x}_{hsg} = 2\Sigma_{i\in G_{hg}}x_{hi}/n_h$. The approximation to v_B given by (7) is then

$$v_B(\hat{T}_R) \doteq \Sigma_h(N_h/n_h)^2(\bar{x}_h/\bar{x}_{hs})^2(d_{1h} - d_{2h}\bar{y}_{hs}/\bar{x}_{hs})^2.$$

This approximate form contains a multiplicative factor based on $\bar{x}_h/\bar{x}_{hs}$ and in that sense is similar to estimators proposed by Royall and Cumberland (1981a) and Wu (1985). Under model (1) we have

$$E_\xi v_B(\hat{T}_R) \doteq \Sigma_h N_h^2 \bar{x}_h^{-2}\{n_h \bar{v}_{hs} + d_{2h}^2 \alpha_h^2/\bar{x}_{hs}^2 + d_{2h}^2 \bar{v}_{hs}/(n_h \bar{x}_{hs}^2) - d_{2h}[\bar{v}_{hs1} - \bar{v}_{hs2}]/\bar{x}_{hs}\}/(n_h^2 \bar{x}_{hs}^{-2}) \quad (8)$$

where $\bar{v}_{hsg} = 2\Sigma_{G_{hg}} v_{hi}/n_h$. If a stratified sample balanced on x is selected, $\hat{T}_R$ is ξ-unbiased under (1) and in that special type of sample, it is reasonable to estimate the prediction-variance. The last three terms in the braces create a large-sample ξ-bias, when estimating $var_\xi(\hat{T}_R - T)$, that can be eliminated by forcing $\bar{x}_{hs1} = \bar{x}_{hs2}$, a condition we will refer to as α-balance. If SRS assignment to groups is used, α-balance will be achieved asymptotically. However, $d_{2h}^2 = O_r(n_h)$ and under mild regularity conditions the four terms in (10) when summed over the strata have orders N^2/n, N^2/n, N^2/n^2, and N^2/n^2. In particular, the term involving α_h is $O_r(N^2/n)$ so that, as an estimator of $var_\xi(\hat{T}_R - T)$, v_B may have a positive ξ-bias in stratified balanced samples if SRS assignment is used.

Of particular interest is the model having $\alpha_h = 0$ and $v_{hi} = \sigma_h^2 x_{hi}$ for which $\hat{T}_R$ is the best linear ξ-unbiased predictor of T. In that case v_B has a modest ξ-bias of order N^2/n^2 if SRS assignment is used. When $\alpha_h = 0$, v_B is also a robust estimator of the ξ-variance under the general variance specification v_{hi}.

4.2 Separate Linear Regression Estimator

The algebra is quite involved for this estimator and we will only sketch the results. The separate regression estimator $\hat{T}_{LR}$ is ξ-unbiased under (1) and taking the expectation of approximation (5) shows that $v_B(\hat{T}_{LR})$ is ξ-unbiased to terms of order N^2/n. Under model (1) α-balance is unnecessary to obtain approximate ξ-unbiasedness for v_B.

5. AN EMPIRICAL STUDY

The theory of the preceding sections was tested in an empirical study using stratified simple random sampling and the estimators of totals described in section 2. Both conditional and unconditional properties were examined and are compared in this section.

The finite population in the study consisted of 1184 iron and steel foundries in the United States. The variable y was the number of employees each establishment had in March 1980 and the auxiliary x was the number of employees one year earlier. Figure 1 is a plot of y versus x for a stratified sample of 200 establishments from the population. Based on the plot both the ratio and the regression estimator appear to be reasonable choices. The coefficients of the simple linear model $\alpha + \beta x$ with $var(y) \propto x$, fitted from the entire finite population, were $\alpha = 2.95$ and $\beta = .92$. Several specifications of the form $var(y) \propto x^\gamma$, $0 \leq \gamma \leq 1$, were tried. All fit the population well with nonzero values of γ producing a somewhat better fit for large y values.

The population was divided into five strata using a method similar to the cumulative square root rule (Cochran 1977, p.129). Establishments were sorted in ascending order on x and approximately equal-sized strata were created based on the cumulative value of $x^{\frac{1}{2}}$. Simple random samples of equal size were selected without replacement from each stratum. This sample selection procedure was repeated 2000 times for two sample sizes — 10 units per stratum for a total of 50 and 40 units per stratum for a total sample of 200. For each sample the separate ratio and regression estimators of the total were computed.

Two versions of the balanced half-sample variance estimator were computed for each estimated total. Within each stratum sample units were assigned to two groups using each of two methods: (a) random assignment and (b) a type of purposive assignment called the basket method (Wallenius 1980). A set of eight orthogonally balanced half-samples was selected from the pairs of groups formed in each stratum. The versions of the BHS estimator will be denoted v_B(ran) and v_B(bas) corresponding to the two methods of assigning sample units to groups.

According to earlier theory, random assignment should produce approximately unbiased estimates of the (p,ξ)-MSE generally and should give satisfactory estimates of the ξ-MSE as long as $\hat{T}$ is ξ-unbiased. The basket method is designed to more nearly meet the α-balance condition $\bar{x}_{hs1} = \bar{x}_{hs2}$, defined for $\hat{T}_R$, in individual samples. The method begins by sorting the sample units from a stratum from high to low based on x. The unit with the largest x is assigned to the first group and the unit with the second largest x to the second group. The unit with the third largest x is assigned to the group with the smaller x-total and so on. When (1) holds, this method may reduce the overestimation by v_B of the ξ-variance of $\hat{T}_R$ in stratified balanced samples, but in samples where $\hat{T}_R$ is ξ-biased the basket method may result in underestimates of the ξ-MSE.

A finite population correction (fpc) factor was also included in the calculation of v_B. For each of the estimators of totals, N_h was replaced by $N_h' = N_h(1-f_h)^{\frac{1}{2}}$ in the calculation of $\hat{T}^{(\alpha)}$ and $\hat{T}$ in (4) — an approach suggested by McCarthy (1966) for estimators which are linear with respect to the p-distribution. The use of N_h' has the effect of inserting the fpc $1-f_h$ into the approximate formula (5) for $\hat{T}_R$ and $\hat{T}_{LR}$. In this study the inclusion of the fpc was critical because f_h ranges from .02 to .14 when n = 50 and from .07 to .57 when n = 200.

The linearization variance estimator defined in expression (6) and the jackknife estimator defined in (7) were also included in the study for comparison. For $\hat{T}_R$ and $\hat{T}_{LR}$ the specific forms of the linearization estimator are respectively :

$$v_L = \Sigma_h N_h^2(1-f_h)\Sigma_{s_h} r_{1hi}^2/[n_h(n_h-1)] \text{ and}$$

$$v_L = \Sigma_h N_h^2(1-f_h)\Sigma_{s_h} r_{2hi}^2/[n_h(n_h-2)] \text{ where}$$

$r_{1hi} = y_{hi} - x_{hi}\bar{y}_{hs}/\bar{x}_{hs}$ and $r_{2hi} = (y_{hi}-\bar{y}_{hs}) - b_{hs}(x_{hi}-\bar{x}_{hs})$.

Table 1 gives summary statistics for the variance estimates over all 2000 samples for each of the two sample sizes. For $\hat{T}_R$ all four choices of variance estimator are nearly unbiased for the empirical p-MSE at either sample size with the exception of v_B(bas) which is on average too small when n=200. For $\hat{T}_{LR}$ v_B(ran) is a severe overestimate when n=50 while v_B(bas) is not, but this disparity disappears at the larger sample size. The estimator v_L is an underestimate at both sample sizes. At the larger sample size the version of the BHS estimator with random assignment is, on average, nearer the p-MSE than the version with purposive assignment.

Table 2 gives 95% confidence coverage results over the 2000 samples based on each estimator $\hat{T}$ and its associated variance estimators. The standardized error (SZE) defined as $(\hat{T}-T)/v^{\frac{1}{2}}$ was computed in each sample for each $\hat{T}$ and accompanying variance estimator v. The percentages of samples with SZE<-1.96, SZE>1.96, and |SZE|≤1.96 were then computed. When |SZE|≤ 1.96 the normal approximation 95% confidence interval covers the population total T. Both BHS estimators give relatively poor coverage rates at either sample size with coverage percentages ranging from 85.6 to 91.8 with v_B(ran) giving somewhat better results than v_B(bas). Note that despite the severe overestimation of the p-MSE of $\hat{T}_{LR}$ by v_B(ran) when n=50, the coverage rate for that case is less than the nominal 95%. The estimators v_L and v_J generally produce somewhat better unconditional coverage percentages than the BHS choices, particularly when n=200.

Important differences not apparent in Tables 1 and 2 emerge among the variance estimators when conditional analyses are performed. The theory sketched earlier showed that under model (1) the ξ-biases of $\hat{T}_R$ and the linearization estimators for $\hat{T}_R$ and $\hat{T}_{LR}$ depend on the degree of within-stratum balance on x, of which $\bar{x}_s = \Sigma_h N_h \bar{x}_{hs}/N$ is an approximate measure. To test this theory the samples were sorted in ascending order by $\bar{x}_s$ and divided into 10 groups of 200 samples each. In each group the averages of $\bar{x}_s$, the error $\hat{T}$-T for each estimator of the total, the empirical root MSE, and the square roots of the averages of the variance estimators were computed. Figures 2 and 3 are trajectory plots of the results.

The figures make it clear that the estimators of totals can be substantially biased in samples which are not near the balance point $\bar{x}_s = \bar{x}$ and that the performance of the variance estimators also depends on $\bar{x}_s$. The bias squared of $\hat{T}_R$ is a substantial part of the MSE in samples where $\bar{x}_s$ is extreme for both n=50 and n=200. $\hat{T}_{LR}$ is more nearly conditionally unbiased but also has problems when $\bar{x}_s$ is extreme. For $\hat{T}_R$, v_B(ran), v_B(bas), and v_J follow the MSE reasonably well at either sample size except when $\bar{x}_s$ is small where all are underestimates. For $\hat{T}_{LR}$ those three variance estimators perform about equally well when n=200, but for n=50 v_B(ran) is a substantial overestimate of MSE($\hat{T}_{LR}$) throughout the range of $\bar{x}_s$. The linearization estimator is the poorest for both $\hat{T}_R$ and $\hat{T}_{LR}$, being a systematic underestimate at either sample size when $\bar{x}_s < \bar{x}$.

Recall from section 4 that v_B, with either method of assignment, is approximately ξ-unbiased under the cases of (1) for which $\hat{T}_R$ and $\hat{T}_{LR}$ are ξ-unbiased. The fact that the BHS estimators track the conditional, empirical MSE's relatively well through much of the range of $\bar{x}_s$, even though the $\hat{T}$'s are conditionally biased is quite similar to the findings of Royall and Cumberland (1981a,b) in unstratified samples. Specifically, variance estimators that are ξ-unbiased under a reasonable prediction model may be fairly robust estimators of the ξ-MSE when the model fails. This contrasts to the case of v_L which is not approximately ξ-unbiased under the models for which $\hat{T}_R$ and $\hat{T}_{LR}$ are. This theoretical deficiency manifests itself in v_L's having poorer conditional performance than the other variance estimators.

To examine conditional confidence interval coverage, the samples were sorted into the same ten groups as above and the percentage of SZE's less than -1.96 and greater than 1.96 were computed in each group. Figures 4 and 5 are similar to those found in Royall and Cumberland (1985) and show the percentages plotted versus the average $\bar{x}_s$ in each group. The top of each bar gives the percentage of SZE's greater than 1.96; the bottom gives the percentage less than -1.96. Horizontal reference lines are drawn at 2.5% above and below zero where the top and bottom of each bar would fall if the distribution of the SZE was standard normal.

Figures 4 and 5 include percentages of extreme SZE's based on v_L, v_J, and v_B(ran). The percentages based on v_B(bas) are similar to, though slightly larger than, those for v_B(ran) for both $\hat{T}_R$ and $\hat{T}_{LR}$ and are not shown. When $\bar{x}_s$ is small each of the variance estimators produces an excessive number of SZE's greater than 1.96 reflecting the positive biases of $\hat{T}_R$ and $\hat{T}_{LR}$ when $\bar{x}_s < \bar{x}$. When $\bar{x}_s$ is large and $\hat{T}_R$ has a negative bias, the opposite occurs — the variance estimators produce an excess of SZE's less than -1.96. The one exception for $\hat{T}_R$ is v_L at n=50 whose positive bias when $\bar{x}_s$ is large helps prevent large negative SZE's. For $\hat{T}_{LR}$ the BHS estimator with random assignment to groups is an improvement over v_L when n=50, but v_L performs better than v_B(ran) when n=200, particularly in samples near the balance point $\bar{x}_s = \bar{x}$. Generally, v_J yields the best conditional coverage percentages, but even its performance suffers when $\bar{x}_s$ is extreme.

The relatively poor performance of v_B for confidence interval construction may be due in part to its instability compared to the other choices studied here. Table 3 gives the square roots of the empirical MSE's of the variance estimators over all 2000 samples. When n=200, for example, the root of the MSE of v_B(ran) is 50% larger than that of v_J for $\hat{T}_R$ and 33% larger than that of v_J for $\hat{T}_{LR}$.

Finally, because v_B(ran) is the standard version of the grouped BHS estimator in the literature, we examined its performance in more depth. From the groups used in Figures 2-5, based on ordering samples by $\bar{x}_s$, the 800 samples in the fourth through the seventh group were selected. In these groups of samples $\hat{T}_R$ and $\hat{T}_{LR}$ were less biased and v_B(ran) should be more nearly ξ-unbiased in large samples. For each sample the quantity $B = \Sigma_h N_h |\bar{x}_{hs1} - \bar{x}_{hs2}|$ was computed after units were randomly assigned to the two groups in each stratum. The 800 samples were then sorted from low to high by B, which is a measure of the degree of α-balance produced by random assignment, and divided into 8 groups of 100 samples each. In each group the ratio $R = [v_B(\text{ran})/\text{MSE}]^{\frac{1}{2}}$ and the percentage of samples with $|\text{SZE}| \leq 1.96$ were computed with the results shown in Table 4. Group 1 contains the samples with the smallest values of B and group 8 the largest. In samples where random assignment produced the largest values of B, R is about 1.3 for $\hat{T}_R$ at both sample sizes and for $\hat{T}_{LR}$ when n=200. When n=50, R for $\hat{T}_{LR}$ ranges from 1.08 in the first group to 2.47 in the eighth group. This implies that if estimation of the ξ-MSE is the goal, then random assignments should be avoided which produce an extreme imbalance on x between the groups. However, this is not the case for confidence interval coverage where the percentage of samples having $|\text{SZE}| \leq 1.96$ is generally nearer the nominal 95% in groups where B is large and v_B(ran) overestimates the MSE.

We conclude this section with two brief asides. A version of v_B using systematic assignment of sample units to the two groups within a stratum after sorting by x was also included in the simulation study. This version of v_B was generally intermediate in performance between v_B(ran) and v_B(bas). Second, we note that the conditional deficiencies of v_L can be largely remedied by adjustments derived from model-based arguments given by Royall and Cumberland (1981a,b). Similar, adjusted versions of v_L have also been studied by Wu (1985).

6. CONCLUSION

The conventional method of applying the grouped version of the balanced half-sample variance estimator v_B is to randomly assign sample units to two groups within each stratum and to select balanced replications from the groups. Other studies have given this method theoretical and empirical support when the goal is estimation of the probability sampling MSE of a variety of linear and nonlinear statistics. This conventional application of v_B also produces estimators of the conditional MSE's of the separate ratio and regression estimators that are theoretically robust in large samples under models with quite general variance specifications as long as the estimators of totals themselves are conditionally unbiased. This theoretical finding was borne out in the empirical study reported here but only so long as samples were not seriously imbalanced on the auxiliary x used in constructing the ratio and regression estimators.

In the empirical examination here the jackknife estimator v_J was generally superior for conditional and unconditional inference to any version of the grouped BHS estimator we considered. This was particularly true for confidence interval construction. Part of the reason for this may be instability of v_B, as noted here and in the earlier theoretical study by Krewski (1978).

The empirical findings here are in some disagreement with others in the literature. Wolter (1985, section 8), for example, summarizes results from two empirical studies involving cluster sampling in which v_B, without grouping of clusters, was better than v_L and v_J for confidence interval construction but was inferior to v_L in terms of bias. Extensions to multistage sampling of the work here will be needed to determine the properties of v_B when pairs within strata are formed by grouping clusters.

REFERENCES

Cochran, W.G. (1977). Sampling Techniques (3rd ed.). New York: John Wiley & Sons.

Cumberland, W.G. and Royall, R.M.(1981),"Prediction models and unequal probability sampling," Journal of the Royal Statistical Society B, 43, 353-367.

Dippo, C.S. (1981)."Variance estimation for nonlinear estimators based upon stratified samples from finite populations," unpublished Ph.D. dissertation, George Washington University.

Dippo C.S. and Wolter, K.M. (1984)."A comparison of variance estimators using the Taylor series approximation," Proceedings of the Section on Survey Research Methods, American Statistical Association, 113-121.

Holt, D. and Smith, T.M.F. (1979),"Post stratification," Journal of the Royal Statistical Society A, 142, 33-46.

Jones, H.L. (1974),"Jackknife estimation of functions of stratum means," Biometrika, 61, 343-348.

Kish, L. and Frankel, M.R. (1970),"Balanced repeated replication for standard errors," Journal of the American Statistical Association, 65, 1071-1094.

Krewski, D. (1978),"On the stability of some replication variance estimators in the linear case," Journal of Statistical Planning and Inference, 2, 45-51.

Krewski, D. and Rao, J.N.K. (1981),"Inference from stratified samples : properties of the linearization, jackknife, and balanced repeated replication methods," Annals of Statistics, 9, 1010-1019.

McCarthy, P.J. (1966),"Replication : an approach to analysis of data from complex surveys," Vital and Health Statistics, Series 2, No.14, U.S. government printing office, Washington, D.C.

——— (1969a),"Pseudoreplication : half-samples," Review of the International Statistical Institute, 37, 239-264.

——— (1969b),"Pseudoreplication : further evaluation and application of the balanced half-sample technique," Vital and Health Statistics, Series 2, No. 31, U.S. government printing office, Washington, D.C.

Rao, J.N.K. and Wu, C.J.F. (1985),"Inference from stratified samples : second order analysis of three methods for nonlinear statistics," Journal of the American Statistical Association, 80, 620-630.

Royall, R.M. (1976),"Current advances in sampling theory : implications for human observational studies," American Journal of Epidemiology, 104, 463-473.

Royall, R.M. and Cumberland, W.G. (1978),"Variance estimation in finite population sampling,"Journal of the American Statistical Association, 73, 351-358.

——— (1981a),"An empirical study of the ratio estimator and estimators of its variance," Journal of the American Statistical Association, 76, 66-77.

——— (1981b),"The finite population linear regression estimator and estimators of its variance — an empirical study,"Journal of the American Statistical Association, 76, 924-930.

——— (1985),"Conditional coverage properties of finite population confidence intervals," Journal of the American Statistical Association, 80, 355-359.

Rust, K. (1985),"Variance estimation for complex estimators in sample surveys," Journal of Official Statistics, 1, 381-397.

Smith, T.M.F. (1984),"Present position and potential developments : some personal views, sample sample surveys," Journal of the Royal Statistical Society A, 147, 208-221.

Valliant, R. (1987a),"Some prediction properties of balanced half-sample variance estimators in single-stage sampling," Journal of the Royal Statistical Society B, 49, accepted for publication.

——— (1987b),"Conditional properties of some estimators in stratified sampling," Journal of the American Statistical Association, 82, accepted for publication.

Wallenius, K. (1980),"Statistical methods in sole source contract negotiation," Journal of Undergraduate Mathematics and its Applications, 0, 35-47.

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

Wu, C.J.F. (1985),"Variance estimation for the combined ratio and combined regression estimators, Journal of the Royal Statistical Society B, 47, 147-154.

Table 1. Summary statistics for estimates of totals and variances from 2000 stratified simple random samples selected from a population of iron and steel foundries.

Summary quantity	Sample size 50	Sample size 200
Separate ratio estimator		
Relative error of $\hat{T}_R$	.004	.001
$MSE^{\frac{1}{2}}$	8.6	3.8
$[\text{Avg. var. est.}/MSE]^{\frac{1}{2}}$		
$v_B(ran)$	1.02	.97
$v_B(bas)$	1.02	.93
v_L	.97	.96
v_J	1.03	.98
Separate regression estimator		
Relative error of $\hat{T}_{LR}$	.009	.002
$MSE^{\frac{1}{2}}$	9.0	3.7
$[\text{Avg. var. est.}/MSE]^{\frac{1}{2}}$		
$v_B(ran)$	1.54	.99
$v_B(bas)$	1.10	.95
v_L	.86	.92
v_J	1.18	1.00

Notes: Relative error of each $\hat{T}$ is computed as $\Sigma(\hat{T}-T)/T$. The MSE is computed as $\Sigma(\hat{T}-T)^2/2000$. Summations are over the 2000 samples. Root MSE's are in thousands.

Table 2. Summary statistics for standardized errors from 2000 stratified simple random samples from a population of iron and steel foundries.

Estimators of the total and variance		Sample size	Percentage of SZE's: SZE<-1.96	|SZE|≤1.96	SZE>1.96
$\hat{T}_R$	$v_B(ran)$	50	6.3	86.3	7.4
		200	6.6	87.5	5.9
	$v_B(bas)$	50	5.5	86.3	8.2
		200	8.2	85.6	6.2
	v_L	50	2.2	91.5	6.3
		200	2.8	93.6	3.6
	v_J	50	2.0	93.7	4.3
		200	2.6	94.4	3.0
$\hat{T}_{LR}$	$v_B(ran)$	50	2.4	91.8	5.8
		200	5.3	89.2	5.5
	$v_B(bas)$	50	3.6	88.3	8.1
		200	6.2	87.9	5.9
	v_L	50	3.6	88.6	7.8
		200	3.5	93.3	3.2
	v_J	50	1.1	94.6	4.3
		200	2.5	94.8	2.7

Table 3. Square roots of empirical mean squared errors of variance estimates over 2000 stratified simple random samples. (Figures are in millions.)

Variance estimator	Sample size 50	Sample size 200
Separate ratio estimator		
$v_B(ran)$	84.5	11.3
$v_B(bas)$	84.7	10.1
v_L	65.0	7.1
v_J	72.8	7.6
Separate regression estimator		
$v_B(ran)$	477.1	12.5
$v_B(bas)$	154.2	10.8
v_L	61.4	7.0
v_J	165.5	9.4

Table 4. Summary statistics for the BHS estimator with random assignment of units to groups for 800 samples with $\bar{x}_s$ near $\bar{x}$.

	Separate ratio estimator				Separate regression estimator			
Group	$R=[\bar{v}_B(ran)/MSE]^{\frac{1}{2}}$		|SZE|≤1.96		$R=[\bar{v}_B(ran)/MSE]^{\frac{1}{2}}$		|SZE|≤1.96	
	n=50	n=200	n=50	n=200	n=50	n=200	n=50	n=200
1	.89	.95	80	85	1.08	.97	85	87
2	1.00	1.00	93	90	1.29	1.00	96	90
3	1.02	1.04	83	94	1.25	1.04	87	93
4	1.00	.94	91	86	1.31	1.01	94	86
5	1.01	.93	87	87	1.56	.97	92	85
6	1.07	1.01	91	87	1.57	.99	95	90
7	1.09	1.19	86	94	1.71	1.15	94	92
8	1.31	1.28	92	93	2.47	1.27	96	95

Notes: Groups were formed based on size of $B=\Sigma_h N_h|\bar{x}_{hs1}-\bar{x}_{hs2}|$ computed for each sample. $\bar{v}_B(ran)=\Sigma v_B(ran)/100$ with the summation over samples in a group. Column for SZE gives percentage of samples in a group with $|SZE|\le 1.96$.

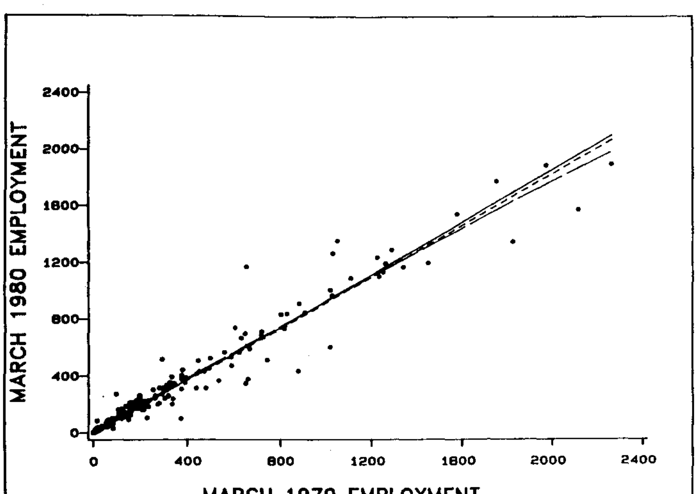

Figure 1. March 1980 employment plotted versus March 1979 employment for iron and steel foundries in the United States. Curves are weighted least squares regression lines under a model with $var(y) \propto x$. Solid line is $E(y)=\beta x$; short-dash line is $E(y)=\beta_0+\beta_1 x$; long-dash line is $E(y)=\beta_0+\beta_1 x+\beta_2 x^2$.

Figure 2. Conditional bias, root mean squared error, and standard error estimates for the separate ratio estimator $\hat{T}_R$ in two sets of 2000 stratified simple random samples. Error curve is average value of $\hat{T}_R$-T; ran, bas, L, and J curves are square roots of averages of v_B(ran), v_B(bas), v_L, and v_J; MSE curve is square root of average $(\hat{T}_R-T)^2$.

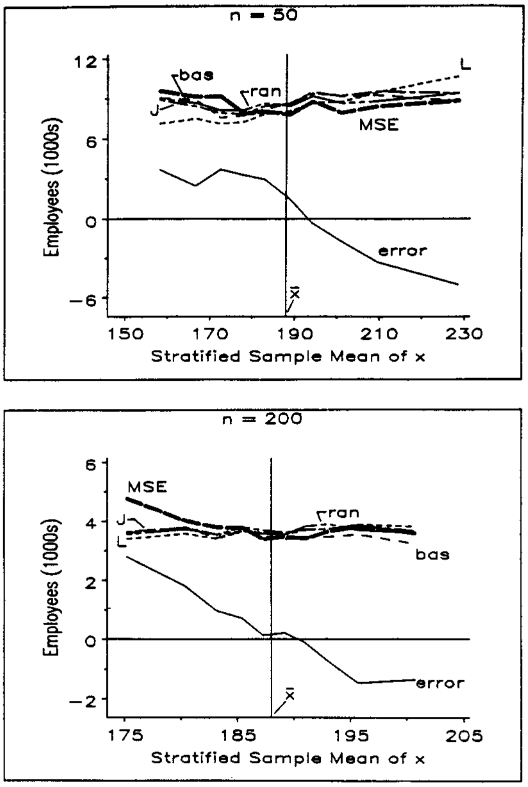

Figure 3. Conditional bias, root mean squared error, and standard error estimates for the separate regression estimator $\hat{T}_{LR}$ in two sets of 2000 stratified simple random samples. Error curve is average value of $\hat{T}_{LR}$-T; ran, bas, L, and J curves are square roots of averages of v_B(ran), v_B(bas), v_L, and v_J; MSE curve is square root of average $(\hat{T}_{LR}-T)^2$.

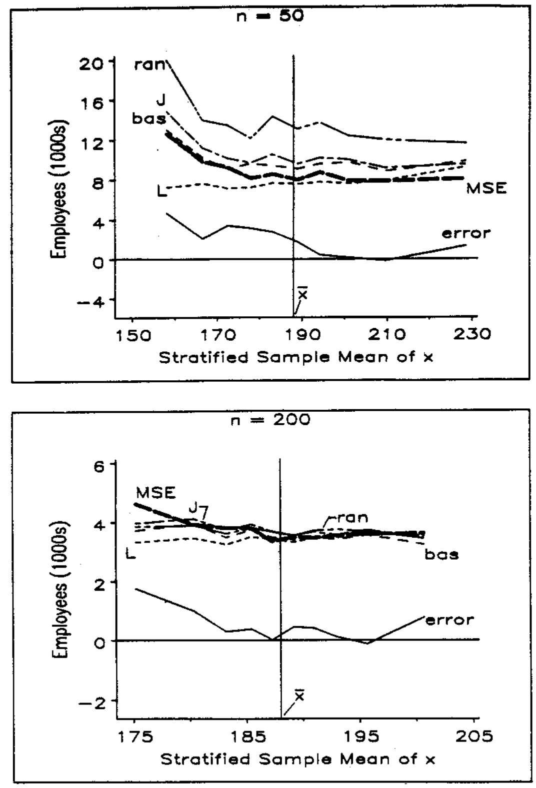

Figure 4. Conditional error percentages of 95% confidence intervals based on $\hat{T}_R$ with v_L, v_J, and v_B(ran) in two sets of 2000 stratified simple random samples. Top of each bar gives percentage of samples with SZE>1.96; bottom gives percentage with SZE<-1.96. Horizontal reference lines are drawn at 2.5 percent above and below zero.

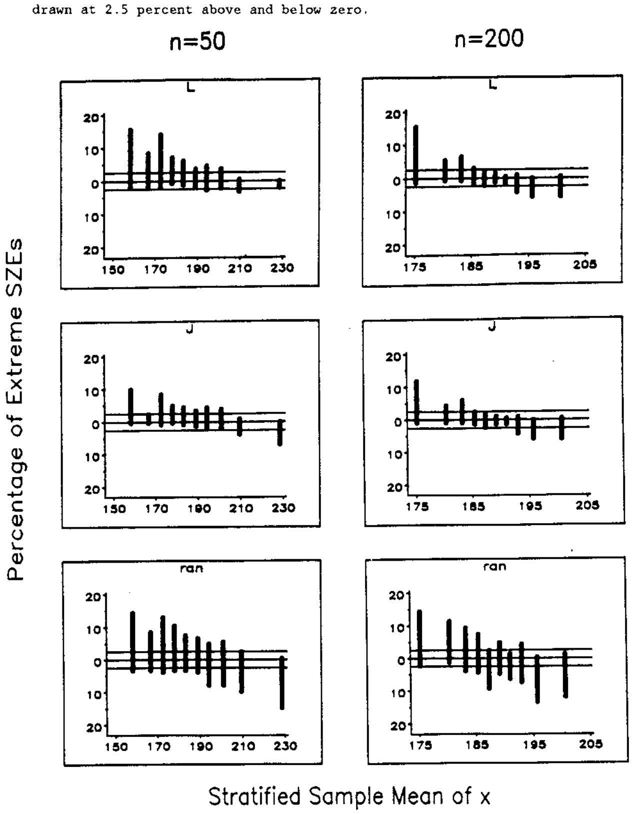

Figure 5. Conditional error percentages of 95% confidence intervals based on $\hat{T}_{LR}$ with v_L, v_J, and v_B(ran) in two sets of 2000 stratified simple random samples. Top of each bar gives percentage of samples with SZE>1.96; bottom gives percentage with SZE<-1.96. Horizontal reference lines are drawn at 2.5 percent above and below zero.