Post-Stratification and Conditional Variance Estimation

Richard Valliant
U.S. Bureau of Labor Statistics
Room 2126 , 441 G St. NW
Washington DC 20212

February 1992

Richard Valliant is a mathematical statistician in the Office of Survey Methods Research, U.S. Bureau of Labor Statistics. He thanks the referees for their useful comments.

ABSTRACT

Post-stratification estimation is a technique used in sample surveys to improve efficiency of estimators. Survey weights are adjusted to force the estimated numbers of units in each of a set of estimation cells to be equal to known population totals. The resulting weights are then used in forming estimates of means or totals of variables collected in the survey. For example, in a household survey the estimation cells may be based on age/race/sex categories of individuals and the known totals may come from the most recent population census. Although the variance of a post-stratified estimator can be computed over all possible sample configurations, inferences made conditionally on the achieved sample configuration are desirable. Theory and a simulation study using data from the U.S. Current Population Survey are presented to study both the conditional bias and variance of the post-stratified estimator of a total. The linearization, balanced repeated replication, and jackknife variance estimators are also examined to determine whether they appropriately estimate the conditional variance.

Keywords: Asymptotic properties; Balanced repeated replication; Jackknife variance estimation; Linearization variance estimation; Superpopulation model.

1. INTRODUCTION

In complex large-scale surveys, particularly household surveys, post-stratification is a commonly used technique for improving efficiency of estimators. A clear description of the method and the rationale for its use was given by Holt and Smith (1979) and is paraphrased here. Values of variables for persons may vary by age, race, sex, and other demographic factors that are unavailable for sample design at the individual level. A population census may, however, provide aggregate information on such variables that can be used at the estimation stage. After sample selection, individual units are classified according to the factors and the known total number of units in the *c*th cell, M_c, is used as a weight to estimate the cell total for some target variable. The cell estimates are then summed to yield an estimate for the full population. A variety of government-sponsored household surveys in the United States use this technique, including the Current Population Survey, the Consumer Expenditure Survey, the National Health Interview Survey, and the Survey of Income and Program Participation.

Because post-stratum identifiers are unavailable at the design stage, the number of sample units selected from each post-stratum is a random variable. Inferences can be made either unconditionally, i.e. across all possible realizations of the post-strata sample sizes, or conditionally given the achieved sample sizes. In a simpler situation than that considered here, Durbin (1969) maintained, on grounds of common sense and the ancillarity of the achieved sample size, that conditioning was appropriate. In the case of post-stratification in conjunction with simple random sampling of units, Holt and Smith (1979) argue strongly that inferences should be conditioned on the achieved post-stratum sample sizes.

Although conditioning is, in principle, a desirable thing to do, a design-based conditional theory for complex surveys may be intractable, as noted by Rao (1985). A useful alternative is the prediction or superpopulation approach which is applied in this paper to make inferences from post-stratified samples. We will concentrate especially on the properties of several commonly used variance estimators to determine whether they estimate the conditional variance of the post-stratified estimator of a finite population total.

Section 2 introduces notation, a superpopulation model that will be used to study properties of various estimators, and a class of estimators which will be used as the starting point for post-stratification estimation. Section 3 discusses the model bias and variance of estimators of the total while sections 4 through 6 cover the linearization, balanced repeated replication, and jackknife variance estimators. In section 7 we present the results of a simulation study using data from the U.S. Current Population Survey and the last section gives concluding remarks.

2. NOTATION AND MODEL

The population of units is divided into H design strata with stratum h containing N_h clusters. Cluster (hi) contains M_{hi} units with the total number of units in stratum h being $M_h = \sum_{i=1}^{N_h} M_{hi}$ and the total in the population being $M = \sum_{h=1}^{H} M_h$. A two-stage sample is selected from each stratum consisting of $n_h \geq 2$ sample clusters and a subsample of m_{hi} sample units within sample cluster (hi). The total number of clusters in the sample is $n = \sum_h n_h$. The set of sample clusters from stratum h is denoted by s_h and the subsample of units within sample cluster (hi) by s_{hi}.

Associated with each unit in the population is a random variable y_{hij} whose finite population total is $T = \sum_h \sum_{i=1}^{N_h} \sum_{j=1}^{M_{hi}} y_{hij}$. Each unit is also a member of a class or post-stratum indexed by c. Each post-stratum can cut across the design strata and the set of all population units in post-stratum c is denoted by S_c. The total number of units in post-stratum c is $M_c = \sum_h \sum_{i=1}^{N_h} \sum_{j=1}^{M_{hi}} \delta_{hijc}$ where $\delta_{hijc} = 1$ if unit (hij) is in post-stratum c and is 0 if not. We assume that the post-stratum sizes M_c are known. Our goal here will be to study the properties of estimators under the following superpopulation model:

$$E(y_{hij}) = \mu_c$$

$$\operatorname{cov}(y_{hij}, y_{h'i'j'}) = \begin{cases} \sigma^2_{hic} & h = h', i = i', j = j', (hij) \in S_c \\ \sigma^2_{hic}\rho_{hic} & h = h', i = i', j \neq j', (hij) \in S_c, (h'i'j') \in S_c \\ \tau_{hicc'} & h = h', i = i', j \neq j', (hij) \in S_c, (h'i'j') \in S_{c'} \\ 0 & \text{otherwise} \end{cases} . \qquad (1)$$

In addition to being uncorrelated, we also assume that the *y*'s associated with units in different clusters are independent. The model assumes that units in a post-stratum have a common mean μ_c and are correlated within a cluster. The size of the covariances $\sigma^2_{hic}\rho_{hic}$ and $\tau_{hicc'}$ are allowed to vary among the clusters and also depend on whether or not units are in the same post-stratum. The variance specification σ^2_{hic} is quite general, depending on the design stratum, cluster, and post-stratum associated with the unit. All expectations in the subsequent development are with respect to model (1) unless otherwise specified.

The general type of estimator of *T* that we will consider has the form

$$\hat{T} = \sum_h \sum_{i \in s_h} \gamma_{hi} \hat{T}_{hi} \qquad (2)$$

where γ_{hi} is a coefficient that does not depend on the *y*'s, $\hat{T}_{hi} = M_{hi}\bar{y}_{hi}$, and $\bar{y}_{hi} = \sum_{j \in s_{hi}} y_{hij} / m_{hi}$. In common survey practice, the set of γ_{hi} is selected to produce a design-unbiased or design-consistent estimator of the total under the particular probability sampling design being used. Alternatively, estimator (2) can be written as

$$\hat{T} = \sum_h \sum_{i \in s_h} \sum_c K_{hic} \bar{y}_{hic} \qquad (3)$$

where $K_{hic} = \gamma_{hi} M_{hi}\, m_{hic} / m_{hi}$, m_{hic} is the number of sample units in sample cluster (*hi*) that are part of post-stratum *c*, and $\bar{y}_{hic} = \sum_{j \in s_{hi}} y_{hij} \delta_{hijc} / m_{hic}$. If $m_{hic} = 0$, then define $\bar{y}_{hic} = 0$. There are a variety of estimators, both from probability sampling theory and superpopulation theory, that fall in this class. Six examples are given in Valliant (1987) and include types of separate ratio and regression estimators with M_{hi} used as the auxiliary variable. For example, the ratio estimator has $\gamma_{hi} = M_h \big/ \sum_{i \in s_h} M_{hi}$ **Error! Reference source not found.**, and the regression estimator has

$$\gamma_{hi} = \frac{N_h}{n_h}\left[1 + n_h\left(\bar{M}_h - \bar{M}_{hs}\right)\left(M_{hi} - \bar{M}_{hs}\right) \Big/ \sum_{i \in s_h}\left(M_{hi} - \bar{M}_{hs}\right)^2\right]$$

where $\bar{M}_h$ and $\bar{M}_{hs}$ are population and sample means per cluster of the M_{hi}'s. Also included in the class defined by (2) is the Horvitz-Thompson estimator when clusters are selected with probabilities proportional to M_{hi} and units within clusters are selected with equal probability in which case $\gamma_{hi} = M_h \big/ \left(n_h M_{hi}\right)$. Note that, as discussed in section 3, the estimators defined by (2) are not necessarily model-unbiased under (1).

Next, we turn to the definition of the post-stratified estimator of the total. The usual design-based estimator of M_c in class (2) is found by using δ_{hijc} in place of y_{hij} in (3) and omitting the sum over *c*, which gives $\hat{M}_c = \sum_h \sum_{i \in s_h} K_{hic}$. The post-stratified estimator of the total *T* is then defined as

$$\hat{T}_{ps} = \sum_c \hat{R}_c \hat{T}_c \qquad (4)$$

where $\hat{R}_c = M_c \big/ \hat{M}_c$ **Error! Reference source not found.** and $\hat{T}_c = \sum_h \sum_{i \in s_h} K_{hic}\bar{y}_{hic}$. With this notation the general estimator (3) can also be written as $\hat{T} = \sum_c \hat{T}_c$. For subsequent calculations it will be convenient to write down the model for the set of means $\{\bar{y}_{hic}, i \in s_h\}$ implied by model (1):

$$E(\bar{y}_{hic}) = \mu_c$$

$$\operatorname{cov}(\bar{y}_{hic}, \bar{y}_{h'i'c'}) = \begin{cases} v_{hic} & h = h', i = i', c = c' \\ v_{hicc'} & h = h', i = i', c \neq c' \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

where $v_{hic} = \sigma^2_{hic}\left[1 + \left(m_{hic} - 1\right)\rho_{hic}\right] \big/ m_{hic}$.

3. MODEL-BIAS AND VARIANCE OF ESTIMATORS OF THE TOTAL

The model bias under (1) of the unadjusted estimator $\hat{T}$ is $E(\hat{T} - T) = \sum_c \mu_c \left(\hat{M}_c - M_c \right)$. Estimators in class (2) are model unbiased if $\hat{M}_c = M_c$, a condition which in general does not hold for a particular sample, but may be true in expectation across all samples that a particular design can produce. For example, when clusters are selected with probability proportional to M_{hi} and an equal probability sample of units is selected within each sample cluster, the Horvitz-Thompson estimator has $\hat{M}_c = \sum_h \sum_{i \in s_h} \left[M_h / \left(n_h m_{hi} \right) \right] \sum_{j \in s_{hi}} \delta_{hijc}$, whose expectation under the design is M_c. On the other hand, the post-stratified estimator $\hat{T}_{ps}$ is model-unbiased under (1), as is easily verified.

In many populations a model requiring a common mean for all units in post-stratum c may be unrealistic. For example, in household surveys post-strata are often based on age, race, and sex while design strata are, in part, based on geography and degree of urbanization. One post-stratum might be white males, aged 35-39, while design strata might be defined based on region of the country and an urban/rural dichotomy. Assuming that white males, aged 35-39 in the urban northeast United States, have the same mean as the same age group of white males in the rural southeast, may be unreasonable. If the correct model has $E(y_{hij}) = \mu_{hc}$ for $(hij) \in S_c$, then the bias of $\hat{T}_{ps}$ is

$$E(\hat{T}_{ps} - T) = \sum_c \sum_{i \in s_h} \mu_{hc} \left(\hat{F}_c \hat{M}_{hc} - M_{hc} \right) \qquad (6)$$

where $\hat{M}_{hc} = \sum_{i \in s_h} K_{hic}$. If the sample is balanced in the sense that the estimated proportion of the post-stratum c units that are within design stratum h is equal to the population proportion ($\hat{M}_{hc} / \hat{M}_c = M_{hc} / M_c$), then $\hat{T}_{ps}$ is model-unbiased, but in many probability samples this will not be the case.

The prediction variance of the post-stratified estimator is defined as $\text{var}\left(\hat{T}_{ps} - T \right)$. Under some reasonable assumptions, similar to those given in Royall (1986) or Valliant (1987), on how certain population and sample quantities grow as $H \to \infty$, we have

$$\operatorname{var}\left(\hat{T}_{ps} - T\right) \approx \operatorname{var}\left(\hat{T}_{ps}\right) \tag{7}$$

where $\approx$ denotes "asymptotically equivalent to." Details are sketched in Appendix A.1. Consequently, we will concentrate on the estimation of $\operatorname{var}(\hat{T}_{ps})$.

In order to compute the variance, it is convenient to write the post-stratified estimator as $\hat{T}_{ps} = \sum_h \hat{\mathbf{R}}'\hat{\mathbf{T}}_h$ where $\hat{\mathbf{R}} = \left(\hat{R}_1, \ldots, \hat{R}_C\right)'$,**Error! Bookmark not defined.** $\hat{\mathbf{T}}_h = \left(\mathbf{K}'_{h1}\bar{\mathbf{y}}_{h1}, \ldots, \mathbf{K}'_{hC}\bar{\mathbf{y}}_{hC}\right)'$, $\mathbf{K}_{hc} = \left(K_{h1c}, \ldots, K_{hn_hc}\right)'$, and $\bar{\mathbf{y}}_{hc} = \left(\bar{y}_{h1c}, \ldots, \bar{y}_{hn_hc}\right)'$. Using the model for the means $\bar{y}_{hic}$ given by (5), the variance can be found, as sketched in the appendix, as

$$\operatorname{var}(\hat{T}_{ps}) = \sum_h \hat{\mathbf{R}}'\mathbf{K}'_h\mathbf{V}_h\mathbf{K}_h\hat{\mathbf{R}} \tag{8}$$

where $\mathbf{K}'_h$ is the $C \times n_hC$ matrix whose c^{th} row is $\left(\mathbf{0}'_{n_h}, \ldots, \mathbf{0}'_{n_h}, \mathbf{K}'_{hc}, \mathbf{0}'_{n_h}, \ldots, \mathbf{0}'_{n_h}\right)$, i.e. $\mathbf{K}'_{hc}$ is preceded by c-1 zero row vectors of length n_h and followed by C-c such zero vectors. The matrix $\mathbf{V}_h$ is defined as

$$\mathbf{V}_h = \begin{bmatrix} \mathbf{V}_{h1} & \mathbf{D}_{\tau h12} & \cdots & \mathbf{D}_{\tau h1C} \\ \mathbf{D}_{\tau h21} & \mathbf{V}_{h2} & & \\ \vdots & & \ddots & \\ \mathbf{D}_{\tau hC1} & & & \mathbf{V}_{hC} \end{bmatrix}$$

with $\mathbf{V}_{hc} = \operatorname{diag}\left(v_{hic}\right)_{n_h \times n_h}$ and $\mathbf{D}_{\tau hcc'} = \operatorname{diag}\left(\tau_{hicc'}\right)_{n_h \times n_h}$ where $i \in s_h$ for both $\mathbf{V}_{hc}$ and $\mathbf{D}_{\tau hcc'}$. A key point is that, although the factors $\hat{R}_c$ may be random with respect to the sample design, they are constant with respect to model (1), so that (8) is a variance conditional on the values of $\hat{R}_c$.

The variance of the unadjusted estimator $\hat{T}$ can be found by minor modification of the above arguments. Because $\hat{T} = \sum_h \hat{\mathbf{T}}_h$, i.e. the value of $\hat{T}_{ps}$ when $\hat{R}_c = 1$ for all c, we have

$$\operatorname{var}(\hat{T}) = \sum_h \mathbf{1}'_C\mathbf{K}'_h\mathbf{V}_h\mathbf{K}_h\mathbf{1}_C \tag{9}$$

where $\mathbf{1}_C$ is a vector of C 1's. Note that, if the sample design is such that the post-stratum factors $\hat{R}_c$ each converge to 1, then $\operatorname{var}\left(\hat{T}_{ps}\right)$ and $\operatorname{var}\left(\hat{T}\right)$ are about the same in large samples.

4. A LINEARIZATION VARIANCE ESTIMATOR

Linearization or Taylor series variance estimators for post-stratified estimators are discussed for general sample designs by Rao (1985) and Williams (1962). An application to a complex survey design is given in Parsons and Casady (1985). Our interest here is in how a linearization estimator, derived from design-based arguments, performs as an estimator of the approximate conditional variance given by (8). For clarity and completeness we will sketch the derivation of the estimator for the class of post-stratified estimators studied here. In a design-based analysis, the product $\hat{R}_c \hat{T}_c$ is expanded about the point (M_c, T_c) where T_c is the finite population total for post-stratum c. The usual first-order Taylor approximation to $\hat{R}_c \hat{T}_c$ is $\hat{R}_c \hat{T}_c \cong T_c + \hat{T}_c - T_c \hat{M}_c / M_c$. From this expression it follows that $\hat{T}_{ps} - T \cong \sum_h \sum_{i \in s_h} \tilde{d}_{hi}$ where $\tilde{d}_{hi} = \sum_{j \in s_{hi}} \sum_c \gamma_{hi} M_{hi} \delta_{hijc} \left[y_{hij} - (T_c / M_c) \right] / m_{hi}$. For computations, the usual procedure is to substitute estimators for the unknown quantities in $\tilde{d}_{hi}$ producing $d_{hi} = \sum_{j \in s_{hi}} \sum_c \gamma_{hi} M_{hi} \delta_{hijc} (y_{hij} - \hat{\mu}_c) / m_{hi} = \sum_c K_{hic} (\bar{y}_{hic} - \hat{\mu}_c)$ where $\hat{\mu}_c = \hat{T}_c / \hat{M}_c$. The linearization variance estimator, including an *ad hoc* finite population correction factor, is then defined as

$$v_L(\hat{T}_{ps}) = \sum_h \frac{n_h}{n_h - 1} (1 - f_h) \sum_{i \in s_h} (d_{hi} - \bar{d}_h)^2 \tag{10}$$

where $f_h = n_h / N_h$ and $\bar{d}_h = \sum_{i \in s_h} d_{hi} / n_h$. Note that, although some post-strata may not be represented in cluster i, the term d_{hi} is still defined as long as s_{hi} is not empty, because each sample unit must be in one of the post-strata.

In order to determine whether the general linearization estimator (10) estimates the conditional variance (8), we examine its large sample behavior. First, write $d_{hi} - \bar{d}_h = \sum_c (d_{hic} - \bar{d}_{hc})$ where $d_{hic} = K_{hic} (\bar{y}_{hic} - \hat{\mu}_c)$ and $\bar{d}_{hc} = \sum_{i \in s_h} d_{hic} / n_h$. Squaring out $(d_{hi} - \bar{d}_h)^2$, assuming f_h is negligible, and using the definition of $\mathbf{v}(\hat{\mathbf{e}}_h)$ in Appendix A.2 gives $v_L(\hat{T}_{ps}) = \sum_h \mathbf{1}_C' \mathbf{v}(\hat{\mathbf{e}}_h) \mathbf{1}_C$. It follows from result (A.2) in the appendix that

$$\frac{n}{M^2}\left[v_L\left(\hat{T}_{ps}\right) - \sum_h \mathbf{1}_C'\mathbf{K}_h'\mathbf{V}_h\mathbf{K}_h\mathbf{1}_C\right] \xrightarrow{p} 0$$

Thus, the linearization estimator v_L actually estimates var$\left(\hat{T}\right)$ given by (9) rather than var$\left(\hat{T}_{ps}\right)$ in (8). In large samples the linearization estimator differs from var$\left(\hat{T}_{ps}\right)$ by a factor that depends on how different the adjustment factors $\hat{\mathbf{R}}$ are from $\mathbf{1}_C$ **Error! Reference source not found.**. This is analogous to the model-bias, observed by Royall and Cumberland (1981), of the linearization variance estimator $v_c = \left(N^2/n\right)(1-f)\sum_{i\in s_h}\left(y_i - x_i\bar{y}_s/\bar{x}_s\right)^2/(n-1)$ for the ratio estimator $\hat{T}_R = N\bar{y}_s\,\bar{x}/\bar{x}_s$ where $\bar{y}_s = \sum_{i\in s} y_i/n$, $\bar{x}_s = \sum_{i\in s} x_i/n$, $\bar{x} = \sum_{i=1}^{N_h} x_i/N$, and s is a set of sample units selected by simple random sampling without replacement.

This conditional bias can be eliminated by using the adjusted deviate $d^*_{hi} = \sum_{j\in s_{hi}}\sum_c \hat{R}_c\gamma_{hi}M_{hi}\delta_{hijc}\left(y_{hij} - \hat{\mu}_c\right)/m_{hi}$ **Error! Reference source not found.** $= \sum_c \hat{R}_c K_{hic}\left(\bar{y}_{hic} - \hat{\mu}_c\right)$ in the linearization estimator. This is similar to adjusting the variance estimator v_c for the ratio estimator by the factor $\left(\bar{x}/\bar{x}_s\right)^2$ to produce the estimator known as v_2 (e.g., see Wu 1982). For later reference define the adjusted linearization estimator using d^*_{hi} as $v^*_L\left(\hat{T}_{ps}\right)$ and note that it can be written as

$$v^*_L\left(\hat{T}_{ps}\right) = \sum_h \frac{n_h}{n_h - 1}\left(1 - f_h\right)\sum_{i\in s_h}\left[\sum_c \hat{R}_c\left(d_{hic} - \bar{d}_{hc}\right)\right]^2 \tag{11}$$

This estimator has also been proposed by Binder (1991) and Rao (1985). A related estimator was also studied by Särndal, Swensson, and Wretman (1989). The fact that this adjusted variance estimator is consistent for var$\left(\hat{T}_{ps} - T\right)$ follows from the large sample equivalence of var$\left(\hat{T}_{ps}\right)$ and var$\left(\hat{T}_{ps} - T\right)$ and from result (A.2) in Appendix A.2.

Example. Post-stratification with simple random sampling without replacement

This is the case studied in detail by Holt and Smith (1979) in which a simple random sample of n units is selected without replacement from a total of N. In this case define $H = 1$, $N_H = N$, $n_H = n$, $\gamma_{Hi} = N/n$, and $M_{Hi} = m_{Hi} = 1$. The general post-stratified estimator (4) is then $\hat{T}_{ps} = \sum_c N_c\bar{y}_c$ where $\bar{y}_c$ is the mean of the sample units in post-stratum c and $N_c = M_c$ is the number of units in the population in post-stratum c. Under a model with $E(y_i) = \mu_c$ and $\text{var}(y_i) = \sigma_c^2$ **Error! Reference source not found.** if unit i is in post-stratum c and with all units uncorrelated, the model variance (8) reduces to $\sum_c N_c^2 \sigma_c^2 / n_c$. The linearization estimator (10), becomes $v_L(\hat{T}_{ps}) = (N^2/n)(1-f)\sum_c (n_c - 1)s_c^2/(n-1)$ where $s_c^2 = \sum_{i \in s_c} (y_i - \bar{y}_c)^2/(n_c - 1)$ and $\bar{y}_c$ is the sample mean for units in post-stratum c. The sample variance s_c^2 **Error! Bookmark not defined.** is a model-unbiased estimator of σ_c^2, and the general form of the approximate model-bias of v_L is $\text{bias}\left[v_L(\hat{T}_{ps})\right] \cong (N^2/n^2)\sum_c (\sigma_c^2/n_c)(n_c^2 - nN_c/N)$. Note that nN_c/N is the expected sample size in a post-stratum under simple random sampling. If, by chance, the allocation to post-strata is proportional, i.e. $n_c/n = N_c/N$, then $v_L(\hat{T}_{ps})$ is model-unbiased but, generally, is not. The adjusted linearization estimator, on the other hand, is approximately conditionally unbiased since $v_L^*(\hat{T}_{ps}) \cong \sum_c N_c^2 s_c^2/n_c$ if n and n_c are large and f is near 0.

5. A Balanced Repeated Replication Variance Estimator

Balanced repeated replication (*BRR*) or balanced half-sample variance estimators, proposed by McCarthy (1969), are often used in complex surveys because of their generality and the ease with which they can be programmed. Suppose $n_h = 2$ in all strata. A set of J half-samples is defined by the indicators

$$\delta_{hi\alpha} = \begin{cases} 1 & \text{if unit } i \text{ is in half-sample } \alpha \\ 0 & \text{if not} \end{cases}$$

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

$$\begin{aligned} \delta_h^{(\alpha)} &= 2\delta_{h1\alpha} - 1 \\ &= \begin{cases} 1 & \text{if unit } h1 \text{ is in half-sample } \alpha \\ -1 & \text{if unit } h2 \text{ is in half-sample } \alpha. \end{cases} \end{aligned}$$

Note also that $-\varsigma_h^{(\alpha)} = 2\varsigma_{h2\alpha} - 1$. A set of half-samples is orthogonally balanced if $\sum_{\alpha=1}^{J} \varsigma_h^{\alpha} = \sum_{\alpha=1}^{J} \varsigma_h^{\alpha}\varsigma_{h'}^{\alpha} = 0 \;\; \forall\, h' \neq h$ with a minimal set of half-samples having $H+1 \le J \le H+4$. One of the choices of balanced half-sample variance estimators is

$$v_{BRR}\left(\hat{T}_{ps}\right) = \sum_{\alpha=1}^{J} \left(\hat{T}_{ps}^{(\alpha)} - \hat{T}_{ps}\right)^2 \Big/ J \qquad (12)$$

where $\hat{T}_{ps}^{(\alpha)} = \sum_c \hat{R}_c^{(\alpha)} \hat{T}_c^{(\alpha)}$ with $\hat{R}_c^{(\alpha)}$ being the post-stratum c adjustment factor and $\hat{T}_c^{(\alpha)}$ being the estimated post-stratum total based on half-sample α, both of which are defined explicitly below. Other, asymptotically equivalent choices involving the complement to each half-sample have been studied by Krewski and Rao (1981) and others, but (12) appears to be the most popular choice in practice.

In applying the *BRR* method, practitioners often repeat each step of the estimation or weighting process, including post-stratification adjustments, for each half-sample. The intuition behind such repetition is that the variance estimator will then incorporate all sources of variability. The goal here is the estimation of a conditional variance. This raises the question of whether, to achieve that goal, the post-stratification factors $\hat{\mathbf{R}}$ should be recomputed for each half-sample or whether the full-sample factors should be used for each half-sample.

First, consider the case in which the factors are recomputed from each half-sample and define $\hat{R}_c^{(\alpha)} = M_c \big/ \hat{M}_c^{(\alpha)}$ to be the factor and $\hat{T}_c^{(\alpha)}$ to be the estimated total for post-stratum c based on half-sample α. In particular, $\hat{T}_c^{(\alpha)} = \sum_h \sum_{i \in s_h} \left(2\varsigma_{hi\alpha} - 1\right) K_{hic}\bar{y}_{hic}$ and $\hat{M}_c^{(\alpha)} = \sum_h \sum_{i \in s_h} \left(2\varsigma_{hi\alpha} - 1\right) K_{hic}$. Next, expand $\hat{R}_c^{(\alpha)}\hat{T}_c^{(\alpha)}$ around the full sample estimates $\hat{R}_c$ and $\hat{T}_c$ to obtain the approximation

$$\begin{aligned} \hat{R}_c^{(\alpha)}\hat{T}_c^{(\alpha)} - \hat{R}_c\hat{T}_c &\cong \hat{R}_c\left(\hat{T}_c^{(\alpha)} - \hat{T}_c\right) - \hat{R}_c\hat{m}_c\left(\hat{M}_c^{(\alpha)} - \hat{M}_c\right) \\ &= \hat{R}_c\left[\sum_h \varsigma_h^{(\alpha)}\left(\Delta_{yhc} - \hat{m}_c\Delta_{Khc}\right)\right] \end{aligned} \qquad (13)$$

where $\Delta_{yhc} = K_{h1c}\bar{y}_{h1c} - K_{h2c}\bar{y}_{h2c}$ and $\Delta_{Khc} = K_{h1c} - K_{h2c}$. The variance estimator (12) can then be approximated as $v_{BRR}\left(\hat{T}_{ps}\right) \cong \sum_{\alpha=1}^{J}\left[\sum_c \hat{R}_c \sum_h \varsigma_h^{(\alpha)} z_{hc}\right]^2$ where $z_{hc} = \Delta_{yhc} - \hat{m}_c\Delta_{Khc}$.

Squaring out the term in brackets and using the fact that $\alpha_h^{(r)2} = 1$ and the orthogonality of $\alpha_h^{(r)}$ and $\alpha_{h'}^{(r)}$ $(h \neq h')$, lead to

$$v_{BRR}\left(\hat{T}_{ps}\right) \cong \sum_h \left(\sum_c \hat{R}_c z_{hc}\right)^2. \quad (14)$$

Squaring out the right-hand side of (14) and noting that $z_{hc} z_{hc'} = n_h \sum_{i \in s_h} \left(d_{hic} - \bar{d}_{hc}\right)\left(d_{hic'} - \bar{d}_{hc'}\right) / \left(n_h - 1\right)$ when $n_h = 2$, it follows that, aside from the factor $1 - f_h$, $v_{BRR}\left(\hat{T}_{ps}\right)$ is approximately equal to the adjusted linearization estimator in (11). Consequently, the *BRR* estimator does appropriately estimate the conditional variance when the number of strata is large and when the post-stratification factors are recomputed for each half-sample.

Suppose, alternatively, that the full-sample factors are used for each half-sample, and denote the resulting estimator as $v_{BRR}^{*}\left(\hat{T}_{ps}\right)$. Expression (13) then becomes $\hat{R}_c^{(r)} \hat{T}_c^{(r)} - \hat{R}_c \hat{T}_c = \hat{R}_c \left(\hat{T}_c^{(r)} - \hat{T}_c\right) = \hat{R}_c \sum_h \alpha_h^{(r)} \Delta_{yhc}$ and the term z_{hc} in (14) reduces to $z_{hc} = \Delta_{yhc}$. By direct calculation the expectation of approximation (14) is

$$E\left[\sum_h \left(\sum_c \hat{R}_c \Delta_{yhc}\right)^2\right] \cong \operatorname{var}\left(\hat{T}_{ps}\right) + \sum_h \boldsymbol{\mu}' \mathbf{V}_{Kh} \boldsymbol{\mu}$$

where $\boldsymbol{\mu}' = \left(\mu_1, \ldots, \mu_C\right)$ and $\mathbf{V}_{Kh}$ is a $C \times C$ matrix with the $(cc')^{th}$ element equal to $\sum_{i \in s_h} \left(K_{hic} - \bar{K}_{hc}\right)\left(K_{hic'} - \bar{K}_{hc'}\right)$ where $\bar{K}_{hc} = \sum_{i \in s_h} K_{hic} / n_h$. Because $\mathbf{V}_{Kh}$ is a type of covariance matrix, it is positive semi-definite. As a result, using the full sample post-stratification factors in each replicate can lead to an overestimate of the variance of $\hat{T}_{ps}$. As will be illustrated in the empirical study in section 7, the overestimation can be severe.

6. A JACKKNIFE VARIANCE ESTIMATOR

Another example of a replication variance estimator is the jackknife which, as defined by Jones (1974), is

$$v_J\left(\hat{T}_{ps}\right) = \sum_h \left(1 - f_h\right) \frac{n_h - 1}{n_h} \sum_{i \in s_h} \left[\hat{T}_{ps(hi)} - \hat{T}_{ps(h)}\right]^2 \quad (15)$$

where $\hat{T}_{ps(hi)}$ is the post-stratified estimator computed after deleting sample cluster (*hi*) and $\hat{T}_{ps(h)} = \sum_{i \in s_h} \hat{T}_{ps(hi)} / n_h$ **Error! Reference source not found.**. Similar to the case of the *BRR* estimator, we can write $\hat{T}_{ps(hi)} = \sum_c \hat{R}_{c(hi)} \hat{T}_{c(hi)}$ where $\hat{R}_{c(hi)} = M_c / \hat{M}_{c(hi)}$ and $\hat{T}_{c(hi)}$ are estimators derived from deleting sample cluster (*hi*). Expanding $\hat{R}_{c(hi)} \hat{T}_{c(hi)}$ around the full sample estimates $\hat{R}_c$ and $\hat{T}_c$ yields

$$\hat{R}_{c(hi)} \hat{T}_{c(hi)} - \hat{R}_c \hat{T}_c \cong \hat{R}_c \left(\hat{T}_{c(hi)} - \hat{T}_c \right) - \hat{R}_c \hat{\mu}_c \left(\hat{M}_{c(hi)} - \hat{M}_c \right). \qquad (16)$$

For purposes of computation, write the estimated class total as $\hat{T}_c = \sum_h N_h \bar{\tilde{y}}_{hc}$ where $\bar{\tilde{y}}_{hc} = \sum_{i \in s_h} \tilde{y}_{hic} / n_h$ and $\tilde{y}_{hic} = f_h K_{hic} \bar{y}_{hic}$ **Error! Reference source not found.**. Defining $\bar{\tilde{y}}_{hc(i)} = \left(n_h \bar{\tilde{y}}_{hc} - \tilde{y}_{hic} \right) / \left(n_h - 1 \right)$ leads to

$$\begin{aligned} \hat{T}_{c(hi)} &= N_h \bar{\tilde{y}}_{hc(i)} + \sum_{h' \neq h} N_{h'} \bar{\tilde{y}}_{h'c} \\ &= \hat{T}_c + \frac{N_h}{n_h - 1} \left(\bar{\tilde{y}}_{hc} - \tilde{y}_{hic} \right) \end{aligned}$$

from which it follows that

$$\hat{T}_{c(hi)} - \hat{T}_{c(h)} = \frac{-n_h}{n_h - 1} \left[K_{hic} \bar{y}_{hic} - \frac{1}{n_h} \sum_{i' \in s_h} K_{hi'c} \bar{y}_{hi'c} \right] \qquad (17)$$

Similarly, $\hat{M}_{c(hi)} - \hat{M}_c = -n_h \left(K_{hic} - \sum_{i' \in s_h} K_{hi'c} / n_h \right) / \left(n_h - 1 \right)$. Substitution of this expression and (17) into (16) and use of the resulting approximation in the formula for the jackknife given by (15) gives

$$v_J \left(\hat{T}_{ps} \right) \cong \sum_h \left(1 - f_h \right) \frac{n_h}{n_h - 1} \sum_{i \in s_h} \left[\sum_c \hat{R}_c \left(d_{hic} - \bar{d}_{hc} \right) \right]^2$$

which is the same as the adjusted linearization variance estimator in (11). Consequently, by the same arguments presented for $v_L^* \left(\hat{T}_{ps} \right)$ **Error! Reference source not found.**, the jackknife estimator is also a consistent estimator of the large sample conditional variance shown in (8).

7. A SIMULATION STUDY

The preceding theory was tested in a simulation study using a fixed, finite population of 10,841 persons who were included in the September 1988 Current Population Survey (CPS). The variables used in the study were weekly wages and hours worked per week for each person. The study population contained 2,826 geographic segments. The segments were those used in the CPS with each being composed of about four neighboring households. Eight post-strata were formed on the basis of age, race, and sex using tabulations of weekly wages on the full population. Table 1 shows the age/race/sex categories which were assigned to each post-stratum, and Table 2 gives the means per person of weekly income and hours worked per week in each post-stratum. As is apparent from Table 2, the means differ considerably among the post-strata, especially for weekly wages.

A two-stage stratified sample design was used in which segments were selected as the first-stage units and persons as the second-stage units. Two sets of 10,000 samples were selected. For the first set, 100 sample segments were selected with probabilities proportional to the number of persons in each segment. For the second set, 200 segments were sampled. In both cases, strata were created to have about the same total number of households and $n_h = 2$ sample segments were selected per stratum. Within each stratum, segments were selected systematically using the method described by Hansen, Hurwitz, and Madow (1953, p. 343). A simple random sample of 4 persons was selected without replacement in each segment having $M_{hi} > 4$. In cases having $M_{hi} \le 4$, all persons in the sample cluster were selected. For the samples of 100 segments, the first-stage sampling fraction was 3.5% (100/2826), and for the samples of 200 segments was 7%.

In each sample, we computed the Horvitz-Thompson estimator $\hat{T}_{HT}$ (which is a special case of the general estimator defined by expression (2)), the post-stratified estimator $\hat{T}_{ps}$, and the five variance estimators v_L, v_L^*, v_{BRR}, v_{BRR}^*, and v_J. For the two *BRR* estimators, the half-sample total $\hat{T}_c^{(\alpha)}$ was computed as $\hat{T}_c^{(\alpha)} = \sum_h \sum_{i \in s_h} \sqrt{1-f_h}\,(2\delta_{hi\alpha} - 1) K_{hic} \bar{y}_{hic}$ which has the effect of inserting finite population correction factors for each stratum in the approximation given by (14). Table 3 presents unconditional results summarized over all 10,000 samples. Empirical mean square errors (*mse*'s) were calculated as $mse\left(\hat{T}\right) = \sum_{s=1}^{S}\left(\hat{T}_s - T\right)^2 \Big/ S$ with S = 10,000 and $\hat{T}$ being either $\hat{T}_{HT}$ or $\hat{T}_{ps}$. Average variance estimates across the samples were computed as $\bar{v} = \sum_{s=1}^{S} v_s / S$ where v_s is one of the five variance estimates considered. The table reports the ratios $\sqrt{\bar{v} \big/ mse\left(\hat{T}\right)}$.

As anticipated by the theory in section 4 the linearization variance estimator v_L is more nearly an estimate of the *mse* of the Horvitz-Thompson estimator $\hat{T}_{HT}$ than of the *mse* of $\hat{T}_{ps}$. In fact, the square root of the average v_L overestimates the empirical square root $mse\left(\hat{T}_{ps}\right)$ from 11% to 17%. The adjusted linearization estimator v_L^*, on the other hand, is approximately unbiased for $mse\left(\hat{T}_{ps}\right)$, as is the jackknife. Of the two *BRR* estimates, the root of the average v_{BRR} performs well while v_{BRR}^* is a serious overestimate as predicted by the theory in section 5. As the sample increases from *n*=100 to *n*=200, the percentage overestimation by $\sqrt{\bar{v}_{BRR}^*}$ drops from 22.9% to 13.3% for wages and from 27.7% to 16.0% for hours. The estimate v_{BRR}^* is also much more variable than either v_L^* or v_{BRR}, as shown in the lower part of Table 3. Based on this study, it is clearly preferable to recompute the post-stratification factors for each half-sample rather than using the full sample factors each time.

The differences in the performance of the variance estimates may not be nearly so pronounced in cases where post-stratification does not result in substantial gains over the Horvitz-Thompson estimator. To illustrate this point, the population was divided into only two post-strata -- males and females. Another set of 10,000 samples was then selected for the case *n*=200. Estimates were made for the variable weekly wages. For this simulation the empirical square root of the *mse* of the Horvitz-Thompson estimator was only 2.9% larger than that of $\hat{T}_{ps}$. This contrasts to the figures in Table 3 for *n*=200 where the root *mse* of $\hat{T}_{HT}$ was 15.7% larger than that of $\hat{T}_{ps}$ for weekly wages. The ratios $\sqrt{\bar{v}/mse(\hat{T}_{ps})}$ for $v_L, v_L^*, v_{BRR}, v_{BRR}^*$, and v_J were respectively 1.026, .998, .963, .989, and .998. None of the variance estimates exhibits any substantial deficiencies, and, in particular, v_L and v_{BRR}^* estimate $mse(\hat{T}_{ps})$ much more closely than in the eight post-stratum simulation reported in Table 3.

Figure 1 is a plot that illustrates the conditional empirical biases of the Horvitz-Thompson and post-stratified estimates for the simulations using eight post-strata. The bias of $\hat{T}_{HT}$ under model (1) can be written as $\sum_c E(T_c)\left[\hat{R}_c^{-1} - 1\right]$ where T_c is the population total for post-stratum *c*. In the unlikely event that $E(T_c)$ is about the same for all post-strata, then $D = \sum_c \left(\hat{R}_c^{-1} - 1\right)$ is proportional to that bias but, more generally, is a measure of sample balance. *D* is also a quantity that can be computed for each sample without knowledge of the expected total in each post-stratum. Samples were sorted in ascending order by the value of *D* and were divided into 20 groups of 500 samples each. The average biases of $\hat{T}_{HT}$ and $\hat{T}_{ps}$ were then computed in each group. The average group biases for wages and hours are plotted in Figures 1 and 2 versus the average group values of *D*. The Horvitz-Thompson estimator has a clear conditional bias even though it is unbiased over all samples while the conditional bias of the post-stratified estimator is much less pronounced. When *D* is extreme, the bias of $\hat{T}_{HT}$ is also a substantial proportion of the root $mse(\hat{T}_{HT})$ at either sample size for both wages and hours.

The square roots of the average variance estimates within each group are also plotted as are the empirical square root *mse*'s for each group. The conditional results are similar to the unconditional ones listed in Table 3. The estimates v_L and v_{BRR}^* are substantial overestimates of $mse(\hat{T}_{ps})$ **Error! Reference source not found.** while the other estimates are approximately unbiased. The estimates v_L^*, v_{BRR}, and v_J were almost the same so that, of the three, only v_L^* is graphed in Figure 1.

The empirical coverage probabilities of 95% confidence intervals were also computed in the eight post-stratum simulations using $\hat{T}_{ps}$ and each of the five variance estimates. The results are what one would expect given the bias characteristics of the variance estimates and are not shown in detail. The three choices v_L^*, v_{BRR}, and v_J each produce intervals with near the nominal coverage probabilities while v_{BRR}^* and v_L produce intervals that cover the population total considerably more often than the nominal 95%.

8. CONCLUSION

Post-stratification is an important estimation tool in sample surveys. Though often thought of as a variance reduction technique, the method also has a role in reducing the conditional bias of the estimator of a total, as illustrated here. The usual linearization variance estimator for the post-stratified total $\hat{T}_{ps}$ actually estimates an unconditional variance as shown here both theoretically and empirically. This deficiency is easily remedied by a simple adjustment which parallels one that can be made for the case of the ratio estimator. Standard application of the *BRR* and jackknife variance estimators does, on the other hand, produce conditionally consistent estimators. An operational question that is sometimes raised in connection with replication estimators is whether to recompute the post-stratification factors for each replicate or to use the full sample factors in each replicate estimate. Judging from the theoretical and empirical results for *BRR* reported here, recalculation for each replicate is by far the preferable course, leading to a variance estimator that is more nearly unbiased and more stable.

An area deserving research, which we have omitted, is the use of post-stratification to correct for sample nonresponse and for coverage problems in sampling frames. In the U.S. Current Population Survey, for example, post-stratification leads to substantial upward adjustments in weights for some demographic groups. For Black males, for example, the upward adjustments can range from less than 1% to more than 35% depending on age group (U.S. Bureau of the Census 1978).

APPENDIX

A.1 Prediction variance of the post-stratified estimator

This appendix sketches the derivation of the prediction variance of the post-stratified estimator of the total and the identification of the dominant term in the variance. Large sample calculations will be done for a situation in which $H \to \infty$, all parameters in model (1) are finite, and in which

(i) $n/M \to 0$,

(ii) $\max_{h,i}\{M_{hi}\}$, $\max_{h,i}\{m_{hi}\}$, $\max_{h}\{N_h\}$ and $\max_{h}\{n_h\}$ are $O(1)$,

(iii) $\max_{h,i,c}\{K_{hic}\} = O(M/n)$

(iv) $\hat{\mathbf{R}} \to \mathbf{R}$, a constant

(v) $\left(n/M^2\right)\sum_h \mathbf{K}_h'\mathbf{V}_h\mathbf{K}_h \to \mathbf{G}$, a $C \times C$ positive definite matrix of constants, and

(vi) $M^{-1}\sum_h E\left|K_{hic}\left(\bar{y}_{hic} - \mu_c\right)\right|^{2+\delta} = O(1)$ for some $\delta > 0$ and for all i and c.

Conditions (i) - (iii) are bounding conditions on sample and population sizes. The three are consistent with surveys having a large number of strata, but with population and sample sizes in any stratum being restricted. Condition (iv) requires that each post-stratum factor $\hat{R}_c$ converge to a constant, which is not necessarily 1, as $H \to \infty$. Condition (v) requires the covariance matrix of $\left(\hat{T}_1, \ldots, \hat{T}_C\right)'$ to have a limit when multiplied by the normalizing factor n/M^2. Finally, (vi) is a standard Liapounov condition on moments of $K_{hic}\bar{y}_{hic}$. Next, consider the prediction error of $\hat{T}_{ps}$ which is

$$
\begin{aligned}
\hat{T}_{ps} - T &= \sum_h \sum_c \sum_{i \in s_h} \sum_{j \in s_{hi}} \left(\hat{R}_c \frac{K_{hic}}{m_{hic}} - 1 \right) y_{hij}\delta_{hijc} \\
&\quad - \sum_h \sum_c \sum_{i \in s_h} \sum_{j \notin s_{hi}} y_{hij}\delta_{hijc} - \sum_h \sum_c \sum_{i \notin s_h} \sum_{j=1}^{M_{hi}} y_{hij}\delta_{hijc} \\
&= A - B - C
\end{aligned}
$$

where A, B, and C are defined by the last equality. Using this decomposition, the prediction variance under model (1) is

$$
\operatorname{var}\left(\hat{T}_{ps} - T\right) = \operatorname{var}(A) + \operatorname{var}(B) + \operatorname{var}(C) - 2\operatorname{cov}(A, B).
$$

First, examine var(A) and define $L_{hic} = \left(\hat{R}_c K_{hic} / m_{hic} - 1\right) m_{hic}$. From assumptions (i) - (iii) above, **Error! Reference source not found.** and under the model for the means $\bar{y}_{hic}$ given by (5), we have

$$\operatorname{var}(A) = \sum_h \sum_{i \in s_h} \left\{ \sum_c L_{hic}^2 v_{hic} + \sum_c \sum_{c' \neq c} L_{hic} L_{hic'} \tau_{hicc'} \right\}$$
$$\approx \sum_h \mathbf{R}' \mathbf{K}_h' \mathbf{V}_h \mathbf{K}_h \mathbf{R} .$$

From (iv) and (v) $\operatorname{var}(A) = O\left(M^2/n\right)$.

In order to compute the other components of the variance, some additional notation will be used. Define $v_{1hic} = \sigma_{hic}^2\left[1 + \left(M_{hic} - m_{hic} - 1\right)\rho_{hic}\right]$ and $v_{2hic} = \sigma_{hic}^2\left[1 + \left(M_{hic} - 1\right)\rho_{hic}\right]$. The other components of the prediction variance under model (1) and their orders of magnitude are then obtained by direct computation as

$$\operatorname{var}(B) = \sum_h \sum_{i \in s_h} \sum_c \left\{ \left(M_{hic} - m_{hic}\right) v_{1hic} + \sum_{c' \neq c} \left(M_{hic'} - m_{hic'}\right) \tau_{hicc'} \right\}$$
$$= O(n) \text{ from (ii),}$$

$$\operatorname{var}(C) = \sum_h \sum_{i \notin s_h} \sum_c \left\{ M_{hic} v_{2hic} + \sum_{c' \neq c} M_{hic} M_{hic'} \tau_{hicc'} \right\}$$
$$= O(N - n) \text{ from (ii), and}$$

$$\operatorname{cov}(A, B) = \sum_h \sum_c \hat{R}_c \sum_{i \in s_h} K_{hic} m_{hic} \left\{ \left(M_{hic} - m_{hic}\right) \sigma_{hic}^2 \rho_{hic} + \sum_{c' \neq c} \left(M_{hic'} - m_{hic'}\right) \tau_{hicc'} \right\}$$
$$= O(M) \text{ from (ii), (iii), and (iv),}$$

Pursuant to these calculations and (i), the dominant term of the prediction variance is $\operatorname{var}(A)$ and it is easy to show that $\operatorname{var}(A) \approx \operatorname{var}\left(\hat{T}_{ps}\right)$.

A.2 Other large sample properties

The conditions listed in section A.1 are also sufficient to derive the large sample distribution of the post-stratified estimator and a property of one of the quantities associated with each of the variance estimators considered in the earlier sections. Two key results are that as $H \to \infty$, if conditions (ii), (iii), (v), and (vi) hold, then

(1) $$\frac{\sqrt{n}}{M}\left(\hat{T}_1 - \mu_1 \sum_h \sum_{i \in s_h} K_{hi1}, \ldots, \hat{T}_C - \mu_C \sum_h \sum_{i \in s_h} K_{hiC}\right)' \xrightarrow{d} N(0, \Gamma), \text{ and} \qquad \text{(A.1)}$$

(2) $$\frac{n}{M^2}\left[\sum_h \mathbf{v}(\hat{\mathbf{T}}_h) - \sum_h \mathbf{K}_h' \mathbf{V}_h \mathbf{K}_h\right] \xrightarrow{p} 0 \qquad \text{(A.2)}$$

where $\mathbf{v}(\hat{\mathbf{T}}_h)$ is a $C \times C$ matrix whose (c,c')th element is

$$\left[\mathbf{v}(\hat{\mathbf{T}}_h)\right]_{cc'} = \frac{n_h}{n_h - 1} \sum_{i \in s_h} (d_{hic} - \bar{d}_{hc})(d_{hic'} - \bar{d}_{hc'})$$

with $d_{hic} = K_{hic}(y_{hic} - \hat{\mu}_c)$ and $\bar{d}_{hc} = \sum_{i \in s_h} d_{hic}/n_h$. These results can be proved using Lemmas 3.1 and 3.2, given in Krewski and Rao (1981), which are a central limit theorem and a law of large numbers for independent, nonidentically distributed random variables. Details of the proofs are lengthy but routine and, for brevity, are not presented here. As an immediate consequence of (A.1), (A.2), and (iv) from section A.1, we have

$$\frac{(\hat{T}_{ps} - T)}{\sqrt{\sum_h \hat{\mathbf{R}}' \mathbf{v}(\hat{\mathbf{T}}_h) \hat{\mathbf{R}}}} \xrightarrow{d} N(0,1)$$

so that the usual normal theory confidence intervals are justified.

REFERENCES

Binder, D. (in press), "Use of Estimating Equations for Interval Estimation from Complex Surveys," *1991 Proceedings of the Section on Survey Research Methods, American Statistical Association*.

Durbin, J. (1969), "Inferential Aspects of the Randomness of Sample Size in Survey Sampling," in *New Developments in Survey Sampling*, N.L. Johnson and H. Smith, Jr. eds, Wiley-Interscience, 629-651.

Hansen, M.H., Hurwitz, W., and Madow, W. (1953), *Sample Survey Methods and Theory* (Vol. 1), New York: John Wiley.

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.

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. (1969), "Pseudo-replication: Half-Samples," *Review of the International Statistical Institute,* **37**, 239-264.

Parsons, V.L. and Casady, R.J. (1985), "Variance Estimation and the Redesigned National Health Interview Survey, *Proceedings of the Section on Survey Research Methodology*, American Statistical Association, 406-411.

Rao, J.N.K. (1985), "Conditional Inference in Survey Sampling," *Survey Methodology*, **11**, 15-31.

Royall, R.M. (1986), "The Prediction Approach to Robust Variance Estimation in Two-Stage Cluster Sampling," *Journal of the American Statistical Association,* **81**, 119-123.

Royall, R.M. and Cumberland, W.G. (1981), "An Empirical Study of the Ratio Estimator and Estimators of its Variance," *Journal of the American Statistical Association,* **76**, 66-77.

Särndal, C.-E., Swensson, B., and Wretman, J. (1989), "The Weighted Residual Technique for the Estimating the Variance of the General Regression Estimator of the Finite Population Total," *Biometrika*, **76**, 527-537.

U.S. Bureau of the Census (1978), *The Current Population Survey: Design and Methodology*, Technical Paper 40, Dept. of Commerce: Washington DC.

Valliant, R. (1987), "Generalized Variance Functions in Stratified Two-Stage Sampling," *Journal of the American Statistical Association*, **82**, 499-508.

Williams, W.H. (1962), "The Variance of an Estimator with Post-Stratified Weighting," *Journal of the American Statistical Association*, **57**, 622-627.

Wu, C.F.J. (1982), "Estimation of the Variance of the Ratio Estimator," *Biometrika*, **69**, 183-189.

FIGURE TITLES

Figure 1. Conditional bias, root mean squared error, and standard error estimates for weekly wages in the simulation using eight post-strata. Two sets of 10,000 two-stage stratified samples were selected. Samples were sorted on a measure of sample balance and divided into 20 groups of 500 samples each in order to compute conditional properties.

Figure 2. Conditional bias, root mean squared error, and standard error estimates for hours worked per week in the simulation using eight post-strata. Two sets of 10,000 two-stage stratified samples were selected. Samples were sorted on a measure of sample balance and divided into 20 groups of 500 samples each in order to compute conditional properties.

Table 1. Assignment of age/race/sex categories to post-strata. Numbers in cells are post-stratum identification numbers (1 to 8).

	Non-Black		Black	
Age	Male	Female	Male	Female
19 & under	1	1	1	1
20-24	2	3	3	3
25-34	5	6	4	4
35-64	7	8	4	4
65 & over	2	3	3	1

Table 2. Means per person in each of the eight post-strata for weekly wages and hours worked per week.

Post-stratum	No. of persons M_c	Weekly wages	Hours worked
1	815	111.1	23.7
2	691	278.7	37.9
3	829	221.7	34.7
4	955	349.7	38.8
5	1,543	455.9	43.5
6	1,262	319.1	37.5
7	2,541	554.2	43.1
8	2,205	326.9	36.4
Total	10,841	372.3	38.3

Table 3. Unconditional summary results over two sets of 10,000 two-stage stratified samples of 100 and 200 segments each. Eight post-strata, defined in *Table 1* were used.

Summary quantity	Weekly wages n=100	Weekly wages n=200	Hours worked n=100	Hours worked n=200
Empirical $\sqrt{mse}\,(\times 10^3)$				
$\hat{T}_{HT}$	156.1	111.1	6.8	4.9
$\hat{T}_{ps}$	138.0	96.0	6.2	4.3
$\sqrt{\bar{v}_L / mse(\hat{T}_{HT})}$	1.034	1.002	1.031	.993
[Avg. var. est./$mse(\hat{T}_{ps})$]$^{1/2}$				
v_L	1.170	1.159	1.122	1.110
v_L^*	1.004	1.013	.999	.998
v_{BRR}	1.061	.993	1.028	.982
v_{BRR}^*	1.229	1.133	1.277	1.160
v_J	1.006	1.014	1.002	.998
Std. dev. of var. est.($\times 10^6$)				
v_L	5616	1935	14.7	5.1
v_L^*	4290	1500	12.2	4.2
v_{BRR}	4941	1455	13.1	4.1
v_{BRR}^*	6900	2022	20.9	6.2
v_J	4326	1503	12.3	4.2