1. **INTRODUCTION**

The subjects of this paper are two relatively unrelated problems in variance estimation. Research into these problems was motivated by their applicability to the demographic surveys conducted by the Census Bureau, but their potential applications are more general. The first problem, which is the subject of Section 2, is the development of a methodology for pairing strata in one PSU per stratum designs, which minimizes the bias of the resulting variance estimator when using a collapsed stratum estimator of variance. The current designs of the Current Population Survey, the National Crime Survey and the American Housing Survey are examples of one PSU per stratum designs.

The second problem, which is the subject of Section 3, is the development of an alternative to the standard unbiased variance estimator for two PSUs per stratum, without replacement designs, that will have greater precision. The current design of the Survey of Income and Program Participation is essentially this type of design.

2. **OBTAINING A COLLAPSING THAT MINIMIZES THE BIAS OF THE COLLAPSED STRATUM VARIANCE ESTIMATOR**

To obtain variance estimators for one PSU per stratum designs, a collapsed stratum variance estimator is generally employed, as explained in Wolter (1985). The first step in using such an estimator is the partitioning, or "collapsing", of the set of all strata into groups of two or more strata. Most commonly, each such group of strata consists of two actual strata, and the discussion in this section will be confined to this special case. The main purpose of this section will be to describe how the collapsing can be obtained in a fashion that in practice appears to be close to optimal in terms of minimizing the bias of the corresponding variance estimator.

We first present the collapsed stratum variance estimator, employing for the most part the notation of Wolter (1985). Consider a population total $Y$ to be estimated by a linear estimator of the form

$$
\hat{Y} = \sum_{h=1}^{L} \hat{Y}_h,
$$

which is assumed to be even, and $\hat{Y}_h$ is an unbiased estimator of the total in the $h$-th stratum. The collapsing results in $G = L/2$ groups of strata, with $g_1$ and $g_2$ denoting the two strata in the $g$-th group. The collapsed stratum variance estimator $\hat{V}(\hat{Y})$ of $V(Y)$, as given in Hansen, Hurwitz and Madow (1953), or Wolter (1985), reduces in the case of two strata per group to

$$
\hat{V}(\hat{Y}) = \sum_{g=1}^{G} \left( \frac{2A_{g2}}{A_{g1} + A_{g2}} \hat{Y}_{g1} - \hat{Y}_{g2} \right)^2,
$$

(2.1)

where $A_{gh}$ is a known measure associated with stratum $gh$ that tends to be well correlated with $Y_{gh}$. Commonly used values of $A_{gh}$, which will be discussed later in this section, include:

(i) $1$ for all $gh$,

(ii) the population of the $gh$-th stratum from the most recent census.

We simplify (2.1) by substituting

$$
k_{g1} = \frac{2A_{g2}}{A_{g1} + A_{g2}}, \quad k_{g2} = \frac{2A_{g1}}{A_{g1} + A_{g2}},
$$

which yields

$$
\hat{V}(\hat{Y}) = \sum_{g} \left( k_{g1} \hat{Y}_{g1} - k_{g2} \hat{Y}_{g2} \right)^2.
$$

(2.2)

Note that $k_{g1} + k_{g2} = 2$.

To obtain an expression for Bias $\hat{V}(\hat{Y})$, we observe that

$$
E[\hat{V}(\hat{Y})] = \sum_{g} \left( V(k_{g1} \hat{Y}_{g1} - k_{g2} \hat{Y}_{g2}) + [E(k_{g1} \hat{Y}_{g1} - k_{g2} \hat{Y}_{g2})]^2 \right)
= \sum_{g} \left( (k_{g1}^2 \hat{Y}_{g1}^2 + k_{g2}^2 \hat{Y}_{g2}^2) + (k_{g1} \hat{Y}_{g1} - k_{g2} \hat{Y}_{g2})^2 \right),
$$

(2.3)

where $\sigma_{gh}^2 = V(Y_{gh})$. Since

$$
V(\hat{Y}) = \sum_{gh} (\sigma_{g1}^2 + \sigma_{g2}^2),
$$

follows that

$$
\text{Bias} \hat{V}(\hat{Y}) = \sum_{gh} (k_{g1}^2 \hat{Y}_{g1}^2 + k_{g2}^2 \hat{Y}_{g2}^2),
$$

(2.4)

We observe the following about (2.4) in the two cases mentioned previously. In case (i), (2.4) reduces to

$$
\text{Bias} \hat{V}(\hat{Y}) = \sum_{g} (\hat{Y}_{g1} - \hat{Y}_{g2})^2,
$$

(2.5)

since $k_{gh} = 1$, while in case (ii) both terms of (2.4) are generally present. However, in case (ii) if $A_{gh}$ and $Y_{gh}$ are well correlated then the second term in (2.4) generally tends to be smaller than in case (i), and disappears altogether if $A_{gh}$ is proportional to $Y_{gh}$.

Also note that if $\sigma_{g1}^2 = \sigma_{g2}^2$ for all $g$, then the first term in (2.4) is nonnegative since $k_{g1} + k_{g2} = 2$, but that in general it is possible for the first term of (2.4), and (2.4) itself to be negative, as is illustrated by examples in Hartley, Rao and Kiefer (1969).

In order to obtain a collapsing that minimizes (2.4), the value of (2.4) must be known for each possible pairing. If (2.4) only involves PSU or stratum totals then such information is assumed known at the time of the most recent census for any characteristic tabulated in the census. (Of course these values generally change between the time of the census and the time that the survey is conducted. This problem will be ignored for now, but returned to at the end of this section.) In case (i), only stratum totals are involved, so that the condition is met. In case (ii) there are several possible approaches. If $A_{gh}$ is sufficiently close to $A_{g2}$ for all $g$, then one might choose to ignore the first term of (2.4). If that is not acceptable, another possibility is to first rewrite (2.4) by replacing $\sigma_{gh}^2$ by $\sigma_{ghw}^2 + \sigma_{ghb}^2$ where $\sigma_{ghw}$, $\sigma_{ghb}$ denote the within and between PSU variance.
respectively for the $gh$-th stratum. Then

$$\text{Bias } \hat{V}(\hat{Y}) = \begin{bmatrix} g_{gh} \left( k^2 - 1 \right) g_{gh}^2 \\ + \left[ \sum_{i=1}^{g} \left( k^2 - 1 \right) g_{gh}^2 \\ + \sum_{i=1}^{g} \left( g_i Y_i - k Y \right)^2 g_{gh}^2 \right] \right].$$

(2.6)

The terms within the second set of brackets in (2.6) meet the requirement of involving only PSU and stratum totals. However, census data alone cannot be used to obtain a value for the term within the first set of brackets, since $g_{gh}^2$ depends on the particular within PSU sampling procedure employed. Instead, an estimator $\hat{g}_{gh}^2$ of $g_{gh}^2$ could be obtained directly from the sample, and the estimator

$$\hat{V}(\hat{Y}) = \hat{V}(\hat{V}) + \hat{g}_{gh} \left( 1 - k g_{gh} \right) \hat{g}_{gh}$$

(2.7)

used in place of $V(Y)$ to estimate $V(Y)$. If $\hat{g}_{gh}^2$ was an unbiased estimator of $g_{gh}^2$, then $\hat{V}(\hat{V})$ would be the terms within the second set of brackets of (2.6). Although unbiased estimators of within PSU variance are not obtainable for the commonly used within PSU sampling procedures that employ systematic sampling, it may be possible to consider the bias of $\hat{g}_{gh}^2$ small enough to be ignored.

Whatever approach is chosen, it is assumed that for any collapsing, the contribution to the bias of the variance estimator from each pair of strata is known and nonnegative, and we turn to the key question of this section: Given the set of $L$ strata, how should they be paired in order to minimize the bias of the variance estimator. In an attempt to answer this question, the problem will be formulated as a mathematical programming problem. First let the constants $c_{ij}$, $i<j$, $i=1, \ldots, L$, denote the contribution to the bias of the variance estimator from the pair of strata $i$ and $j$. For example, if the bias is given by (2.5), then $c_{ij} = (Y_i - Y_j)^2$. The total bias of the variance estimator corresponding to any collapsing would be

$$\sum_{i=1}^{L} \sum_{j=i+1}^{L} c_{ij} x_{ij},$$

(2.8)

where

$$x_{ij} = 1, \text{ if strata } i \text{ and } j \text{ are paired together},$$

$$= 0, \text{ otherwise}. $$

Then minimizing the bias of the variance estimator is equivalent to minimizing (2.8) subject to the constraints

$$x_{ij} = 0 \text{ or } 1 \text{ for all } i, j, i<j, (2.9)$$

and that for each $i$ exactly one member of the sequence

$$x_{1i} x_{i2} \ldots x_{i(L-1)i} x_{i(i+1)j} x_{i(i+2)(i+3)} \ldots x_{iL}$$

is equal to 1, or equivalently,

$$x_{1i} = 1, \forall i,j, (2.10)$$

and that for each $i$ exactly one member of the sequence

$$x_{ij} = 1, \forall i=1, \ldots, L.$$

(2.11)

The problem defined by (2.8 - 2.10) is an integer programming problem. If $L$ is sufficiently small, an optimal solution can be obtained by using any standard software for solving integer programming problems. Unfortunately the solution time for such problems increases rapidly with increasing $L$, and if $L$ is fairly large it would be impractical to solve the problem in this fashion.

It would be desirable if this integer programming problem could be transformed into a different form of mathematical programming problem that would be more efficient computationally. To this end, we define $c_{ij} = M$ if $i,j$ and $c_{ij} = M$ for each $i$, where $M$ is a suitably large constant, as will be explained later. We then seek to minimize

$$\sum_{1 \leq i < j \leq L} c_{ij} x_{ij},$$

(2.11)

subject to the constraints

$$\sum_{j=1}^{L} x_{ij} = 1, i=1, \ldots, L, (2.12)$$

$$\sum_{i=1}^{L} x_{ij} = 1, j=1, \ldots, L, (2.13)$$

$$x_{ij} = 0 \text{ or } 1, \forall i, j = 1, \ldots, L. (2.14)$$

The problem (2.11 - 2.14) is an assignment problem. Software exists for solving assignment problems in reasonable time even for quite large $L$. The key question is whether an optimal solution to the assignment problem (2.11 - 2.14) leads to an optimal solution to the original integer programming problem (2.8 - 2.10). The answer would be yes if the following conditions were true for an optimal solution to this assignment problem:

$$x_{ij} = 0, i = 1, \ldots, L, \text{ or } (2.15)$$

$$x_{ij} = x_{ji}, i,j = 1, \ldots, L, \text{ or } (2.16)$$

For, if these conditions were satisfied, then as a result of the symmetry in both the $c_{ij}$'s and $x_{ij}$'s, the subset of the optimal $x_{ij}$'s for the assignment problem for which $i,j$ would satisfy (2.10) and the corresponding value of (2.8) would be 1/2 the value of (2.11). Furthermore, the set $x_{ij}$, $i,j$ minimizes (2.8) subject to (2.9), (2.10), since if $x_{ij}$, $i,j$ also satisfied (2.9), (2.10) and if we let $x_{ij} = x_{ji}$ for $i,j, x_{ij}=0$, then the entire set of $x_{ij}$'s would satisfy (2.12 - 2.14) with

$$x_{ij}^* c_{ij} x_{ij}^* = \frac{1}{2} x_{ij}^* c_{ij} x_{ij}^* > \frac{1}{2} x_{ij}^* c_{ij} x_{ij}^*$$

Thus the value of (2.8) for $x_{ij}^*$, $i,j$ is not less than (2.8) for the set $x_{ij}, i,j$.

Now (2.15) always holds if $M$ is set sufficiently large. For example, any $M > L^{\max \{ i,j \}}$ would certainly suffice.

One might believe that (2.16) also always holds since $c_{ij}=c_{ij}$ for all $i,j$. However, this is false, as is established by the following counterexample. Let $L=6$ and take

$$c_{ij} = 0, if i,j \leq 3 \text{ or } i,j > 4, i \neq j,$$

that is $c_{ij}=0$ for all elements of the array in the upper left or lower right quadrants of the array that are not on the main diagonal. Then the following set of $x_{ij}$'s satisfies (2.12 - 2.14) and yields a value of 0 for (2.11),

$$x_{ij} = 0, i,j \leq 3 \text{ or } i,j > 4, i \neq j.$$
and hence is an optimal solution to the assignment problem:

\[ x_{12} = x_{23} = x_{31} = x_{45} = x_{56} = x_{64} = 1, \]

\[ x_{ij} = 0 \text{ for all other } i,j. \]

Clearly this solution does not satisfy (2.16). Nor are there any other feasible solutions to this problem for which (2.16) holds and (2.11) is 0. To see why, observe that if a set of \( x_{ij} \)'s is a feasible solution to (2.11 - 2.14) and if \( x_{12} = x_{21} = 1 \) then \( x_{3j} = 1 \) for \( j = 4, 5 \) or 6 and hence (2.11) would be positive. Similarly, any other feasible solution to this assignment problem for which \( x_{ij} = x_{ji} = 1 \) for some \( i,j \) with \( c_{ij} = 0 \) immediately forces \( x_{ij} = x_{ji} = 1 \) for some \( i,j \) for which \( c_{ij} = 1 \).

Although an optimal solution to (2.11 - 2.14) does not in general lead to an optimal solution to (2.8 - 2.10), it is believed that a nearly optimal solution can generally be obtained in an efficient manner by combining both of these problems as follows. First obtain an optimal solution to the assignment problem and let \( S_1 \) denote the set of strata for which \( x_{ij} = x_{ji} = 1 \), while the set of all remaining strata is denoted by \( S_2 \). The pairing for the strata in \( S_1 \) is defined by this optimal solution to the assignment problem, that is, the 1-\( u \)th and 2-\( u \)th strata are paired if \( x_{1u} = x_{2u} = 1 \). If \( S_2 \) is sufficiently small then the elements in \( S_2 \) can be paired by obtaining an optimal solution to a problem like (2.8 - 2.10), but with \( S_2 \) now viewed as the set of all strata. If \( S_2 \) is still too large for this purpose, it can be partitioned into a collection of say \( t \) subsets \( S_{21}, \ldots, S_{2t} \), such that each such subset \( S_{2k} \) contains an even number of elements; each \( S_{2k} \) is small enough to efficiently obtain a solution to (2.8 - 2.10) with \( S_{2k} \) viewed as the set of all strata; and strata \( i \) and \( j \) are in the same \( S_{2k} \) if either \( x_{ij} = x_{ji} = 1 \) or \( x_{ji} = x_{ij} = 1 \) in the optimal solution to the assignment problem, provided this last requirement can be met without any of the \( S_{2k} \) becoming too large. (The rationale for grouping strata \( i,j \) for which either \( x_{ij} = x_{ji} = 1 \) or \( x_{ji} = x_{ij} = 1 \) in the same \( S_{2k} \) is that such a grouping tends to put together pairs of strata which would have a small contribution to the bias of the variance estimator.) The elements of \( S_{2k} \) are then paired by the optimal solution of (2.8 - 2.10) restricted to \( S_{2k} \).

The procedure just described results in an optimal pairing of strata in \( S_1 \) and either an optimal pairing of strata in \( S_2 \) or, if \( S_2 \) is partitioned, an optimal pairing of strata in each of the \( S_{2k} \)'s. However, it is not necessarily an optimal pairing for the entire set of \( L \) strata, since such a pairing may require that a stratum in one subset be paired with a stratum in another. Although it does not in general yield an optimal solution, it is believed that this approach would provide a good approximate solution in an efficient manner.

Remark: All of the preceding work has been with respect to a single characteristic \( Y \). Since, as a practical matter, the same collapsing would generally be used for variance estimates for all characteristics, the collapsing criteria could be taken to be the minimization of the weighted average of the biases of the variance estimator for several key characteristics, instead of the bias of a single characteristic, that is,

\[ W_k \text{ Bias } \hat{V}(\hat{Y}_k), \tag{2.17} \]

where \( \hat{Y}_k \) is an unbiased estimator of \( Y_k \). If all of these characteristics are considered of equal importance then \( W_k \) would be some value that would serve as a scaling factor. (One possible scaling factor is presented in the example below.) If some variables are more important than others, then \( W_k \) could be taken to be something greater than the corresponding scaling factor for the more important variables and less than the scaling factor for the less important variables.

**Illustrative Example**

The present design of the Current Population Survey (CPS) is used to illustrate this work. This survey has one PSU per stratum design with 379 nonself-representing strata. (There are also self-representing strata that are subject to a collapsed strata procedure since there is no between PSU variance for such strata.) Because \( L \) is odd, one stratum was arbitrarily dropped for this illustration. After the remaining 378 strata are paired, the discarded stratum could then be grouped with one of the 189 pairs resulting in 188 pairs and one group of three strata. The pair that this strata is grouped with could be chosen to minimize the bias of the total collapsing by computing the bias for each of the 189 possible such groupings that could be created and choosing the grouping with smallest bias.

The comparison criterion is the value of (2.17) where the eight characteristics used were:

- **Unemployed, Total**
- **Black**
- **Hispanic**
- **Teenage (16-19)**
- **Civilian Labor Force, Total**
- **Black**
- **Hispanic**
- **Agriculture Employment**

To obtain \( W_k \), a random pairing was first selected and then for each \( k \), \( W_k \) was taken to be \( (1/8) \text{ Bias } \hat{V}(\hat{Y}_k) \) corresponding to the random pairing. The minimization of the objective function with this \( W_k \) amounts to obtaining a particular pairing for which the average, over the eight characteristics, of the ratio of the bias for this pairing to the bias for the random pairing is minimized.

For each \( k \), \( \text{Bias } \hat{V}(\hat{Y}_k) \) was computed separately for the cases (i) and (ii), defined earlier, both for obtaining \( W_k \) and then for computing the objective function. For case (i), (2.5) was of course used in this computation while for case (ii), the second term only of (2.4) was used to obtain \( \text{Bias } \hat{V}(\hat{Y}_k) \). 1980 census data was used for all computations. In case (i), the procedure resulted in sets \( S_1 \) and \( S_2 \) containing 316 and 62 strata respectively. \( S_1 \) was partitioned into 3 subsets, \( S_{11}, S_{12}, S_{13} \) consisting of 18, 20 and 24 strata. In case (ii), \( S_1 \) and \( S_2 \) contained 278 and 100 strata respectively. \( S_2 \) was partitioned in case (ii) into 4 sets of 26, 24 and 24 strata. The assignment problems were solved with software written by James Fagan, while the Sperry Functional Mathematical Programming System was used to solve the integer programming problems.

The value of the objective function (2.8) corresponding to the final pairing obtained from this procedure for each case is presented in the first column of numbers in Table 2.1. The numbers in columns 2-4 provide an indication of the effectiveness of this procedure. Each value in the second column is...
1/2 the corresponding minimum value of the assignment problem (2.11 - 2.14), which is a lower bound on the minimum value for the integer programming problem (2.8 - 2.10). The numbers in the third column are the values of (2.8) corresponding to a pairing by strata size, that is, with the strata ordered in increasing size based on 1980 population, and the smallest stratum paired with the next smallest stratum, etc. The fourth column presents the values of (2.8) averaged over 10 random pairings independent of the random pairing used in computing the W_k's. The fact that this number is reasonably close to 1 in both cases provides an indication that results similar to these in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to these in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to these in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's. The table indicates that the procedure described in this section yields, for this set of data, a pairing with a variance estimator for the pairings used in this time span that roughly averages 10 years away from the 1980 population, and the fact that this number is reasonably close to 1 in both cases provides an indication that results similar to those in this table would be expected if some other random pairing had been used to compute the W_k's.
variability associated with the estimation of the within PSU variance. This class of estimators has the form

$$
\hat{\sigma}_{ij}^2 = \sum_{i,j} \frac{Y_{ij}^2}{\pi_i \pi_j} - \sum_{i,j} \frac{n}{\pi_i \pi_j} \left( \frac{Y_{ij}}{\pi_i} - \frac{Y_{ij}}{\pi_j} \right)^2 + \sum_{i,j} \frac{n}{\pi_i \pi_j} \left( \frac{\sigma^2}{\pi_i} \right)
$$

(3.3)

where the $k_{ij}$'s are constants. (It is understood that $i=j$ is excluded from the second summation in (3.3) and in all other expressions in this section.) Note that (3.2) is a special case of (3.3) with $k_{ij} = \pi_i / (n-1)$, and that in general $k_{ij} \neq k_{ji}$. Furthermore, in order for (3.3) to be an unbiased estimator of (3.1) restrictions must be placed on the $k_{ij}$'s. To establish what these restrictions are, note that the expected value of (3.3) conditioned on the specific set of sample PSUs is

$$
E[\hat{\sigma}_{ij}^2] = \sum_{i,j} \frac{\pi_i}{\pi_j} \left( \frac{Y_{ij}}{\pi_i} - \frac{Y_{ij}}{\pi_j} \right)^2 + \sum_{i,j} \frac{n}{\pi_i \pi_j} \left( \frac{\sigma^2}{\pi_i} \right)
$$

(3.4)

(3.4) can be rewritten in the alternate form

$$
E[\hat{\sigma}_{ij}^2] = \sum_{i,j} \frac{\pi_i}{\pi_j} \left( \frac{Y_{ij}}{\pi_i} - \frac{Y_{ij}}{\pi_j} \right)^2 + \sum_{i,j} \pi_{ij} \left( d_{ij} + k_{ij} \right) \frac{\sigma^2}{\pi_j}
$$

(3.5)

Comparison of (3.5) with (3.1) shows that (3.3) is an unbiased estimator of $V(\hat{\gamma})$ if and only if

$$
\sum_{j} \pi_{ij} (d_{ij} + k_{ij}) = \pi_i, \quad i = 1, \ldots, N.
$$

(3.6)

Furthermore, since by the proof in Raj (1968, Theorem 6.3)

$$
\sum_{j} \pi_{ij} d_{ij} = \pi_i - \pi_i^2, \quad i = 1, \ldots, N,
$$

(3.7)

(3.6) can be rewritten in the alternate form

$$
\sum_{j} \pi_{ij} k_{ij} = \pi_i, \quad i = 1, \ldots, N.
$$

(3.8)

(3.8) is clearly satisfied by $k_{ij} = \pi_i / (n-1)$, the $(d_{ij} + k_{ij})$'s can be quite variable for fixed $i$ with this set of $k_{ij}$'s because of the variability in the $d_{ij}$'s. An alternate set of $k_{ij}$'s which clearly satisfies (3.6) and completely removes the variability of the $(d_{ij} + k_{ij})$'s is given by

$$
k_{ij} = \frac{1}{n-1} - d_{ij}.
$$

(3.9)

However, since $d_{ij}$ can exceed $1/(n-1)$, (3.9) can be negative for some $i,j$'s and negative estimates of variance can result. To avoid this possibility, a second set of constraints on the $k_{ij}$'s

$$
k_{ij} > 0, \quad i, j = 1, \ldots, N, \quad i \neq j
$$

(3.10)

is added to (3.6), and the set of $k_{ij}$'s defined by (3.9) will not be considered further in unmodified form.

One method of modifying (3.9) to satisfy (3.10) is to let

$$
k_{ij} = \frac{1}{n-1} - d_{ij} \quad \text{if } d_{ij} < \frac{1}{n-1},
$$

(3.11)

$$
k_{ij} = 0 \quad \text{otherwise}.
$$

(3.11)

This method was suggested by Robert Fay of the Bureau of the Census (personal communication). However, this set of $k_{ij}$'s does not in general satisfy (3.6) and consequently yields a biased variance estimator.

From now on, we consider only the case where $n=2$ and present what is the major focus of this section, a set of $k_{ij}$'s which satisfies (3.6), (3.10) and which for each $i$ minimizes the deviation of $d_{ij} + k_{ij}$ from 1 in the sense that for each $p > 1$,

$$
E(|d_{ij} + k_{ij} - 1|^{p}) \leq 0, \quad i \text{ sample is in sample}
$$

(3.12)

is minimized subject to (3.6), (3.10). (The expectation in (3.12) is with respect to the other sample PSU. Deviations from 1 are considered because it follows from (3.6) that for fixed $i$ this is the expected value of $d_{ii} + k_{ii}$ given that the $i$-th PSU is in sample. To define this set of $k_{ij}$'s for fixed $i$, we first relabel the sequence $d_{ii}, \ldots, d_{(i-1)i}, d_{(i+1)i}, \ldots, d_{ii}$ to transform it into a nondecreasing sequence. Then let

$$
\alpha_{ij} = \frac{\pi_i}{\pi_j}, \quad i, j = 1, \ldots, N, \quad i \neq j.
$$

(3.13)

Next, let $m_i$ be the largest integer for which $\alpha_{ij} < \alpha_{i(m_i)}$ and finally let

$$
k_{ij} = \alpha_{i(m_i)} - d_{ij} \quad \text{if } j < m_i,
$$

(3.14)

$$
k_{ij} = 0 \quad \text{otherwise}.
$$

(3.14)

Roughly, the motivation for (3.11) is that for each $i$, $d_{ij} + k_{ij}$ becomes a constant function of $j$ except for those $j$ which would require a negative $k_{ij}$ to accomplish this. In fact, if $d_{ij} < 1$, it can be shown that $m_i = 1$ for all $i$, it can be shown that $m_i = 1$ for all $i$, $\alpha_{ij} = N$ for $i \neq N$, $m_i = N-1$, and (3.14) then reduces to (3.9) with $n=2$.

In the complete paper, which is available from the authors, we establish that the $k_{ij}$'s satisfy the stated conditions, that is (3.6) and (3.10) are satisfied and (3.12) is minimized subject to these constraints. This is omitted here due to lack space.

Illustrative Example

We will compare numerically our variance estimator, defined by (3.3), (3.14), with two other estimators previously described, the method given in Raj (1968) and defined by (3.2), and the estimator suggested by Fay and defined by (3.3), (3.11). These three variance estimators will be referred to as the conditional unbiased (CU), unconditional unbiased (U), and conditional biased (CB) estimators respectively. ("Conditional" indicates that $k_{ij}$ is conditioned on $j$.)

The survey used in the comparison was the original 1980 census based design for the Survey of Income and Program Participation (SIPP). (A sample cut took place before this design was ever implemented in which some sample PSUs were dropped, but this cut is not considered here.) There were 95 strata in that design from which two PSUs were selected without replacement. There were also 91 self-representing strata and eight nonself-representing strata from which one PSU was selected per stratum which will not
be considered in this example.

The comparison criterion will be one component of the squared error of (3.3), namely the MSE of the second term in (3.4), which we denote by \( W \), that is

\[
W = (d_{ij} + k_{ij}) - \frac{\hat{\sigma}_i^2}{\pi_i^2} + (d_{ij} + k_{ji}) = \frac{\hat{\sigma}_j^2}{\pi_j^2},
\]

where \( i \) and \( j \) are the sample PSUs. To simplify our computations, it will be assumed that \( \hat{\sigma}_i^2 \) is proportional to \( \pi_i^2 \). Furthermore, since the comparison would not be affected by the constant of proportionality, we take \( \hat{\sigma}_i^2 / \pi_i^2 = 1 \) for all \( i \), and thus \( W \) reduces to

\[
W = 2d_{ij} + k_{ij} + k_{ji}.
\]

Now from (3.6) it follows that

\[
E(W) = \sum_{i,j} \pi_{ij} = 2
\]

for the CU and UU procedures, which is also the value for the second term in (3.2). For the CB procedure we have

\[
E(W) = \sum_{i,j} \pi_{ij} \max \{ d_{ij}, 1 \}.
\]

Furthermore, for all three procedures

\[
V(W) = \sum_{i,j} \pi_{ij} (2d_{ij} + k_{ij} + k_{ji})^2 - E(W)^2.
\]

In addition, for the CB procedure only

\[
\text{Bias } W = E(W) - 2,
\]

while Bias \( W = 0 \) for the other two procedures.

One modification of this work was necessary. In the actual selection of PSUs for SIPP, some small PSUs were combined to form a "rotation cluster" in 18 of the strata. In computing the joint probabilities, the cluster was initially treated as a single PSU. If the cluster was selected, then at any time during the life of the design one of the PSUs in the cluster would be in sample with probability proportional to size. (This was done because a new sample is chosen from the sample PSUs each year. For small PSUs there is not enough distinct ultimate sampling units to last the life of the design. PSUs in a cluster can be rotated in and out of sample to avoid this problem. See Alexander, Ernst and Haas (1982) for more details.) As a result of this procedure \( \pi_{ij} = 0 \) if PSUs \( i \) and \( j \) are both in the rotation cluster, and unbiased estimators of variance are no longer possible. To obtain a class of estimators constructed with the goal of being approximately unbiased, the following modifications were made to (3.3) and (3.6). Let \( T = \{ (i,j) \mid i \text{ or } j \text{ are not in the rotation cluster} \} \), \( T_1 = \{ (i,j) \in T \} \),

\[
N = \sum_{i,j} (\pi_i \pi_j - \pi_{ij}) / \pi_{ij} \sum_{i,j} (\pi_i \pi_j - \pi_{ij}) / \pi_{ij} \in T_1 \}
\]

and \( d^*_{ij} = f d_{ij} (\pi_i, \pi_j) / \pi_{ij} \in T_1 \), Then modify (3.3), (3.6), by substituting \( d^*_{ij} \) for \( d_{ij} \) in these expressions, and only summing over \( j \in T_1 \). (The \( f \) factor is to compensate

for the fact that the modified first term in (3.3) is a summation only over \( (i,j) \in T_1 \). These same substitutions in (3.11) and (3.14) are used to modify the CB and CU procedures. As for the UU procedure, \( k_{ij} = \pi_i \) would not satisfy the modified (3.6), since

\[
\sum_{j} \pi_{ij} d^*_{ij} = \pi_i - \pi_i^2
\]

in general. Instead, take

\[
\pi_i - \sum_{j \in T_1} \pi_{ij} d^*_{ij}
\]

\[
W_{ij} = \frac{-f d_{ij}}{\pi_{ij}} \in T_1 \}
\]

It should also be noted that for some \( i \) it is possible that \( \sum_{j \in T_1} \pi_{ij} d^*_{ij} > \pi_i \), in which case no nonnegative set of \( k_{ij} \) could satisfy the modified (3.6). In particular (3.24) would be negative and CU would not be defined since \( d^*_{ij} > a_{ij} \) for all \( j \in T_1 \).

This problem arose in only 1 of the 95 strata under consideration in this illustration and that stratum was dropped from the example.

For each of the remaining 94 strata, \( V(W) \) was computed for all three methods and the resulting values summed over the 94 strata to obtain the first column of Table 3.1. Similarly, for the UB procedure, Bias \( W \) was computed for each stratum with the sum given in column 2 of this table. Finally, MSE, that is the sum of column 1 and the square of column 2, is presented for each of these three procedures in column 3.

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Variance</th>
<th>Bias</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CU</td>
<td>11.6168</td>
<td>0</td>
<td>11.6168</td>
</tr>
<tr>
<td>UU</td>
<td>20.2374</td>
<td>0</td>
<td>20.2374</td>
</tr>
<tr>
<td>CB</td>
<td>8.2359</td>
<td>4.8941</td>
<td>32.1877</td>
</tr>
</tbody>
</table>

Thus for this particular design, MSE is smallest for the CU procedure.

REFERENCES


405