BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION
RESEARCH REPORT SERIES
No. RR-90/10

SIMULTANEOUS SELECTION OF PRIMARY SAMPLING UNITS FOR TWO DESIGNS

by

Lawrence R. Ernst
U.S. Bureau of the Census
Statistical Research Division
Washington, D.C. 20233

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Report issued: March 15, 1993 (Revised)

Abstract: It is demonstrated, using transportation theory, that controlled selection can be used to solve the following sampling problem. Sample primary sampling units (PSUs) are to be selected with probability proportional to size for two designs, both one PSU per stratum, denoted as D_1 and D_2. The universe of PSUs is the same for each design but the stratifications are different. The goal of the problem is to simultaneously select the sample PSUs for the two designs in a manner which maximizes the expected number of PSUs that are in both samples. This procedure differs from previous overlap procedures in that it yields a better overlap, but is only applicable when the two samples can be selected simultaneously. An important special case occurs when the probability of selection for each PSU in D_1 does not exceed its probability of selection in D_2. The procedure can then guarantee that the D_1 sample PSUs are a subset of the D_2 sample PSUs. A proposed, but since cancelled, expansion of the Current Population Survey would have been a potential application of this special case, which is discussed. Variance formulas for estimators of total under the controlled selection procedure are also presented.

Key words: Controlled selection; Current Population Survey; maximizes; overlap; stratification.

1. Introduction

Consider the following sampling problem. Primary sampling units (PSUs) are to be selected for two designs, denoted as D_1 and D_2, both of which are one-PSU-per-stratum designs. The selection of sample PSUs for each design is to be with probability proportional to a measure of size which need not be the same for the two designs. The universe of PSUs is the same for both designs, but each is stratified independently. The sample PSUs in D_1 are required to be a subset of the sample PSUs in D_2. This necessitates the following assumption:

> The probability of selection for each PSU in D_1 does not exceed the probability of selection of that PSU in D_2. (1.1)

In this paper we demonstrate how the two-dimensional controlled selection procedure of Causey, Cox and Ernst (1985) can be used to satisfy all the conditions of this problem, that is:

> There is one sample PSU selected from each D_1 and D_2 stratum, with the required probabilities. (1.2)

> Each D_1 sample PSU is a D_2 sample PSU. (1.3)

A particular application of this procedure to a proposed expansion of the Current Population Survey (CPS), which motivated this work, is presented in Section 6. Plans for this expansion have since been dropped for budgetary reasons. Some readers may wish to read the beginning of Section 6 before proceeding further, to obtain an understanding of this motivation.

After developing this procedure, this author became aware of a more general result by Pruhs (1989), who considers the same problem without the assumption (1.1), in which case (1.2) and (1.3) cannot, in general, be satisfied simultaneously. Instead, using a graph theory approach, Pruhs presented an algorithm for which (1.2) is satisfied and the following additional condition holds:

> The expected value for the number of sample PSUs common to the two designs is maximized and the actual number in common for any sample is always greater than the expected value minus one. (1.4)

Thus, Pruhs viewed the problem as one of maximizing the number of sample PSUs common to the designs when the sample PSUs are chosen for the two designs simultaneously. Previously, Causey, Cox and Ernst (1985), and Ernst (1986) presented optimal linear programming procedures for maximizing the number of sample PSUs in common to two designs when the two sets of sample PSUs are chosen sequentially. In general, choosing the two samples simultaneously permits a larger expected overlap, but in many applications it is not possible to select the samples simultaneously, such as when the two designs are for the same periodic survey, but the second design is a redesign of the first design done at a later date.

It is shown here that the problem considered by Pruhs can also be solved by the controlled selection procedure of Causey, Cox and Ernst (1985). This approach has two advantages over Pruhs' approach. The controlled selection approach involves solving a sequence of transportation problems. Software is readily available which can solve transportation problems, and the remainder of the controlled selection algorithm is easily programmable. In addition, the proof that the controlled selection procedure satisfies the required conditions is not difficult. In contrast, both the theory and the task of programming the algorithm with Pruhs' graph theory approach appears to be much more complex.

In Section 2, a brief review of the procedure of Causey, Cox and Ernst is given. In Section 3, the formulation of the stated sampling problem as a controlled selection problem is presented. The presentation will first be for the more general problem in which (1.1) is not assumed. It will then be shown, quite simply, that with assumption (1.1), a special case of the general problem arises for which (1.3) is satisfied. In Section 4, methods for avoiding some difficulties caused by rounding error in using this procedure are described, together with an illustrative example. In Section 5, formulas for the between PSUs variances for both designs for the usual estimator of total corresponding to probability proportional to size sampling are presented for the controlled selection procedure. Finally, in Section 6, an application of the procedure to the proposed expansion of the CPS is considered, which includes an empirical comparison, for each design, of between PSUs variances for controlled selection and three other procedures for selecting the sample PSUs for the expanded sample.

2. Review of Controlled Rounding and Controlled Selection Concepts

The concepts of zero-restricted controlled rounding and controlled selection are briefly reviewed here. The reader is referred to Cox and Ernst (1982), and Causey, Cox and Ernst (1985) for more details and motivation on this subject, and for other references.

An $(m+1)\times(n+1)$ array, $A=(a_{ij})$, is said to be a tabular array if

$$\sum_{i=1}^{m} a_{ij} = a_{(m+1)j}, \qquad j=1,\ldots,n+1,$$

$$\sum_{j=1}^{n} a_{ij} = a_{i(n+1)}, \qquad i=1,\ldots,m+1.$$

Such an array can be represented in the form

$$\begin{array}{ccccc|c}
a_{11} & \cdot & \cdot & \cdot & a_{1n} & a_{1(n+1)} \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
a_{m1} & \cdot & \cdot & \cdot & a_{mn} & a_{m(n+1)} \\
\hline
a_{(m+1)1} & \cdot & \cdot & \cdot & a_{(m+1)n} & a_{(m+1)(n+1)}
\end{array}$$

with the internal, row total, column total and grand total cells clear from this diagram.

A zero-restricted controlled rounding of an $(m+1)\times(n+1)$ tabular array, $A=(a_{ij})$, to a positive integer base b is an $(m+1)\times(n+1)$ tabular array, $R(A) = (r_{ij})$, for which

$$r_{ij} = \lfloor a_{ij}/b \rfloor b \text{ or } \lceil a_{ij}/b \rceil b \text{ for all i,j,}$$

where $\lfloor x \rfloor$, $\lceil x \rceil$ denote the greatest integer not exceeding x and the smallest integer not less than x, respectively. If no base is stated, base 1 is understood.

By modeling the controlled rounding problem as a transportation problem, Cox and Ernst (1982) obtained a constructive proof that a zero-restricted controlled rounding exists for every two-dimensional tabular array.

If $S=(s_{ij})$ is an $(m+1)\times(n+1)$ tabular array, a solution to the controlled selection problem S is a finite sequence of arrays, $N_1 = (n_{ij1})$, $N_2 = (n_{ij2})$,..., $N_l = (n_{ijl})$, and associated probabilities, $p_1,...,p_l$, satisfying:

N_k is a zero-restricted controlled rounding of S for all k, (2.1)

$$\sum_{k=1}^{l} p_k = 1, \tag{2.2}$$

$$E(n_{ijk} \mid i,j) = \sum_{k=1}^{l} n_{ijk}\, p_k = s_{ij}, \quad i=1,...,m+1, \quad j=1,...,n+1. \tag{2.3}$$

If S arises from a sampling problem for which s_{ij} is the expected number of sampling units selected in each cell, and the actual number selected in each cell is determined by choosing one of the N_k's with its associated probability, then by (2.1) the deviation of s_{ij} from the number of sampling units actually selected from cell (i,j) is less than 1, whether (i,j) is an internal cell or a total cell. By (2.3) the expected number of sampling units selected is s_{ij}.

In Causey, Cox and Ernst (1985), a solution to the controlled selection problem is obtained by computing the sequences $N_1,...,N_l$ and $p_1,...,p_l$ as follows. For fixed k, to obtain N_k, p_k, begin with the $(m+1)\times(n+1)$ tabular array $A_k = (a_{ijk})$, which is computed as described below. Then N_k is simply a zero-restricted controlled rounding of A_k. To define p_k, first let

$$d_k = \max\{\,|\,n_{ijk} - a_{ijk}|: i=1,...,m+1,\ j=1,...,n+1\}, \tag{2.4}$$

and then let

$$\begin{aligned} p_k &= 1 - d_k \quad \text{if } k=1 \\ &= (1-\sum_{i=1}^{k-1} p_i)\,(1-d_k) \ \text{ if } k>1. \end{aligned} \tag{2.5}$$

Finally, to compute A_k, let $A_1 = S$ and for $k>1$, obtain A_k recursively from $A_{k\text{-}1}$, $N_{k\text{-}1}$, $d_{k\text{-}1}$ by letting for all i,j,

$$a_{ijk} = n_{ij(k\text{-}1)} + (a_{ij(k\text{-}1)} - n_{ij(k\text{-}1)})/d_{k\text{-}1}\,. \tag{2.6}$$

It is shown in Causey, Cox and Ernst (1985) that there is an integer l for which $d_l = 0$ and that this terminates the algorithm; that is $N_1,...,N_l$ and $p_1,...,p_l$ satisfy (2.1)-(2.3).

3. The Controlled Selection Procedure for Selection of Sample PSUs

The procedure begins by construction of an $(m+1)\times(n+1)$ tabular array, S, for which a sequence of arrays, $N_1,...,N_l$, and associated probabilities, $p_1,...,p_l$, satisfying (2.1)-(2.3) lead to a solution of the problem described in the Introduction. To construct S, let m', n' denote the number of strata in D_1 and D_2, respectively, and let $m=m'+1$, $n=n'+1$. Let G_1, G_2 denote the random sets consisting of all sample PSUs in D_1 and D_2, respectively. For $i=1,...,m'$, $j=1,...,n'$, let t_{ij} denote the number of PSUs in both the i-th D_1 stratum and j-th D_2 stratum; let B_{iju} denote the u-th such PSU, $u=1,...,t_{ij}$; and let T denote the set of all triples (i,j,u). For $(i,j,u)\in T$, let $P_{iju\alpha}=P(B_{iju}\in G_\alpha)$, $\alpha=1,2$, and let $P_{iju3} = \min\{P_{iju1}, P_{iju2}\}$. Finally, for $i=1,...m'$, $j=1,...,n'$, let

$$s_{ij} = \sum_{u=1}^{t_{ij}} P_{iju3}, \tag{3.1}$$

$$s_{mj} = 1 - \sum_{i=1}^{m'} \sum_{u=1}^{t_{ij}} P_{iju3}, \tag{3.2}$$

$$s_{in} = 1 - \sum_{j=1}^{n'} \sum_{u=1}^{t_{ij}} P_{iju3}, \tag{3.3}$$

$$s_{mn} = 0, \tag{3.4}$$

and let $S=(s_{ij})$ denote the $(m+1)\times(n+1)$ tabular array with internal elements defined by (3.1)-(3.4). Note that the marginal values for S are as follows:

$$s_{i(n+1)} = 1, \quad i=1,\ldots,m', \quad s_{(m+1)j} = 1, \quad j=1,\ldots,n', \tag{3.5}$$

$$s_{m(n+1)} = n' - \sum_{(i,j,u)\in T} P_{iju3}, \tag{3.6}$$

$$s_{(m+1)n} = m' - \sum_{(i,j,u)\in T} P_{iju3}, \tag{3.7}$$

$$s_{(m+1)(n+1)} = m' + n' - \sum_{(i,j,u)\in T} P_{iju3}. \tag{3.8}$$

Interpretation of the array S will now be provided. For $i=1,\ldots,m'$, $j=1,\ldots,n'$, s_{ij} is the probability that a PSU in the i-th D_1 stratum and j-th D_2 stratum is in $G_1 \cap G_2$, while s_{mj} is the probability that a PSU in the j-th D_2 stratum is in $G_2 \sim G_1$, and s_{in} is the probability that a PSU in the i-th D_1 stratum is in $G_1 \sim G_2$. Thus, cells (i,j) for which $i \le m'$, $j \le n'$ correspond to the selection of PSUs that are in sample for both designs, while internal cells in row m correspond to the PSUs in $G_2 \sim G_1$, and similarly, internal cells in column n correspond to PSUs in $G_1 \sim G_2$.

As for the marginals (3.5)-(3.8), (3.5) arises because D_1 and D_2 are one-PSU-per-stratum designs. (3.6) is the expected number of PSUs in $G_2 \sim G_1$, with an analogous interpretation for (3.7). (3.8) is the expected number of PSUs in $G_1 \cup G_2$.

After computing, using the controlled selection algorithm described in Section 2, a set of arrays, N_k, and associated probabilities, p_k, $k=1,...,l$, satisfying (2.1)-(2.3), the selection of the sample PSUs for the two designs is a two-step process. First one of the N_k's is selected. The internal cells of N_k are either 0 or 1. A 1 in a cell (i,j) with $i \leq m'$, $j \leq n'$, indicates $B_{iju} \in G_1 \cap G_2$ for a single $u=1,...,t_{ij}$. Among the t_{ij} such PSUs, one is selected at the second step with conditional probability

$$P(B_{iju} \in G_1 \cap G_2 \,|\, n_{ijk}=1) = P_{iju3}/s_{ij}, \quad u=1,...,t_{ij}. \tag{3.9}$$

A 1 in a cell (m,j), $j=1,...,n'$, indicates that the PSU selected for G_2 from the j-th D_2 stratum is not in G_1. Among the $\sum_{i=1}^{m'} t_{ij}$ PSUs in the j-th D_2 stratum, one is selected at the second step with conditional probability

$$P(B_{iju} \in G_2 \sim G_1 \,|\, n_{mjk}=1) = (P_{iju2} - P_{iju3})/s_{mj}, \quad i=1,...,m', \quad u=1,...,t_{ij}. \tag{3.10}$$

An analogous expression holds for a 1 in an internal cell in column n.

This two-step procedure just described satisfies (1.2) and (1.4). To establish (1.2), first note that clearly, by (3.5), there is exactly one PSU from each D_i stratum in G_i, i=1,2. To show that each PSU is selected into the G_1 and G_2 samples with the correct probabilities, observe that by (2.3), (3.9) and (3.10), it follows that for each $(i,j,u) \in T$,

$$P(B_{iju} \in G_1 \cap G_2) = P(n_{ijk}=1)\ P(B_{iju} \in G_1 \cap G_2 \,|\, n_{ijk}=1) = P_{iju3},$$

$$P(B_{iju} \in G_2 \sim G_1) = P(n_{mjk}=1)\ P(B_{iju} \in G_2 \sim G_1 \,|\, n_{mjk}=1) = P_{iju2} - P_{iju3}.$$

Consequently, $P(B_{iju} \in G_2) = P_{iju2}$. Similarly, it can be shown that $P(B_{iju} \in G_1) = P_{iju1}$. Hence, (1.2) holds.

To establish (1.4), first note that for any selection procedure satisfying (1.2),

$$P(B_{iju} \in G_1 \cap G_2) \leq P_{iju3}, \quad (i,j,u) \in T,$$

and hence

$$E[\text{card } (G_1 \cap G_2)] \leq \sum_{(i,j,u) \in T} P_{iju3}.$$

Then (1.4) follows, since for the controlled selection procedure, (2.3) and (3.1) yield

$$E[\text{card}\ (G_1 \cap G_2)] = \sum_{i=1}^{m'} \sum_{j=1}^{n'} E(n_{ijk} \mid i,j) = \sum_{(i,j,u) \in T} P_{iju3},$$

and (2.1), (3.5), (3.6) yield

$$\text{card}\ (G_1 \cap G_2 \mid N_k) = \sum_{i=1}^{m'} \sum_{j=1}^{n'} n_{ijk} = n' - n_{m(n+1)} > \sum_{(i,j,u) \in T} P_{iju3} - 1, \quad k=1,\ldots,l.$$

Finally, to show (1.3) holds for this procedure with the additonal assumption (1.1), simply observe that if $P_{iju1} \leq P_{iju2}$ for all $(i,j,u) \in T$, then by (3.3),

$$s_{in} = 1 - \sum_{j=1}^{n'} \sum_{u=1}^{t_{ij}} P_{iju1} = 0, \quad i=1,\ldots,m',$$

and hence $G_1 \sim G_2 = \emptyset$ for all G_1, G_2. Note that in this case, the n-th column can be omitted in defining S. If this is done and n is redefined to be n', then (3.1)-(3.8) would remain unchanged with the exception of (3.3), (3.4) and (3.7) which would no longer be defined.

4. Overcoming Rounding Error Problems

In implementing the procedure described in Section 3, some programming difficulties relating to rounding error arose in the solution of the controlled selection problem (2.1)-(2.3) which subsequently were resolved by adding some additional steps to the procedure. Since the solutions to these problems are not obvious, they are presented here along with an example illustrating relevant portions of the procedure.

As background, the input to the software that performs the zero-restricted controlled roundings must be an integer array, which can be rounded to any positive integer base. To use this software to obtain a zero-restricted controlled rounding of a real-valued array, $A_k=(a_{ijk})$, to the base 1, proceed as follows. Express (a_{ijk}) in the form (a'_{ijk}/b_k) where a'_{ijk} is an integer for each i,j,k and b_k is a positive integer. Obtain a zero-restricted controlled rounding $N'_k=(n'_{ijk})$ of $A'_k=(a'_{ijk})$ to the base b_k. Then $N_k=(n'_{ijk}/b_k)$ is a zero-restricted controlled rounding of A_k to the base 1.

For example in Figure 1, rounding the array $S=A_1$ to obtain N_1 would be the first step in solving the controlled selection problem S. (This example is for the special case described in the last paragraph of Section 3. Here the internal cells in rows 1-3 and row 4 correspond to (3.1) and (3.2), respectively; the row and column marginals with value 1.000 correspond to (3.5); the row 4 marginal to (3.6); and the grand total to (3.8)). Instead of directly rounding A_1 to the base 1, each element in A_1 is first multiplied by b_1=10,000 to obtain an integer array, $A'_1=(a'_{ijk})$, whose internal cells are less than 10,000. (In general b_1 is the smallest integer power of 10 that yields an integer array, A'_1.) Next round A'_1 to the base b_1 to obtain an array $N'_1=(n'_{ijk})$, whose internal elements are 0 or 10,000. Then divide by 10,000 to obtain N_1 in Figure 1.

$$S=A_1=\begin{array}{ccccc|c}
0.0000 & 0.0000 & 0.2067 & 0.7933 & 0.0000 & 1.0000 \\
0.3528 & 0.1922 & 0.0000 & 0.0000 & 0.4550 & 1.0000 \\
0.4224 & 0.2645 & 0.3131 & 0.0000 & 0.0000 & 1.0000 \\
0.2248 & 0.5433 & 0.4802 & 0.2067 & 0.5450 & 2.0000 \\
\hline
1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 & 5.0000
\end{array}$$

$$N_1 = \begin{array}{ccccc|c}
0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 & 0 & 2 \\
\hline
1 & 1 & 1 & 1 & 1 & 5
\end{array}$$

Figure 1. Rounding of $S=A_1$

The software that obtains the zero-restricted controlled rounding N'_k of A'_k uses only integer arithmetic, and consequently no rounding error can occur during this step of the procedure. However, two other types of rounding problems did occur for this example. First, when S was initially computed in the illustrative example, the array S in Figure 1 was not obtained, but instead, due to rounding error in the computation of the cell entries, the array S^*, presented in Figure 2, was obtained.

$$S^* = \begin{array}{ccccc|c}
0.0000 & 0.0000 & 0.2067 & 0.7933 & 0.0000 & 1.0000 \\
0.3528 & 0.1923 & 0.0000 & 0.0000 & 0.4550 & 1.0001 \\
0.4225 & 0.2645 & 0.3131 & 0.0000 & 0.0000 & 1.0001 \\
0.2248 & 0.5433 & 0.4802 & 0.2067 & 0.5450 & 2.0000 \\
\hline
1.0001 & 1.0001 & 1.0000 & 1.0000 & 1.0000 & 5.0002
\end{array}$$

$$S^{**} = \begin{array}{ccccc|c} 0.000000 & 0.000000 & 0.206717 & 0.793282 & 0.000000 & .999999 \\ 0.352783 & 0.192262 & 0.000000 & 0.000000 & 0.454954 & .999999 \\ 0.422456 & 0.264462 & 0.313082 & 0.000000 & 0.000000 & 1.000000 \\ 0.224759 & 0.543275 & 0.480201 & 0.206717 & 0.545046 & 1.999998 \\ \hline 0.999998 & 0.999999 & 1.000000 & 0.999999 & 1.000000 & 4.999996 \end{array}$$

$$N^{***} = \begin{array}{rrrrr|r} 0 & 0 & 206700 & 793300 & 0 & 1000000 \\ 352800 & 192200 & 0 & 0 & 455000 & 1000000 \\ 422400 & 264500 & 313100 & 0 & 0 & 1000000 \\ 224800 & 543300 & 480200 & 206700 & 545000 & 2000000 \\ \hline 1000000 & 1000000 & 1000000 & 1000000 & 1000000 & 5000000 \end{array}$$

Figure 2. Overcoming Rounding Error in Computation of S

The fact that not all row and column totals in S^* have only 0's after the decimal point can result in failure to meet the requirement that there be exactly one PSU in G_i from each D_i stratum, i=1,2. For example, if S^* was used to initiate the controlled selection process, and no further rounding errors occurred in the procedure, then the second row total of 1.0001 would result in two PSUs being selected into G_1 from that D_1 stratum with probability .0001.

To convert from S^* to S and thus avoid this problem, proceed as follows. First a new array S^{**} in Figure 2 is obtained by carrying 6 decimal places instead of 4, with the 6th place truncated. The reason for the choice of 2 extra places for this example is that, since there are fewer than 100 internal cells, this insures that all marginals, including the grand total, will be within .0001 of the desired values. The truncation in the last place insures that none of the marginals exceed their desired values.

Next multiply S^{**} by 10^6 to obtain an integer-valued array, S^{***}, which is omitted from Figure 2. Then obtain a zero-restricted controlled rounding N^{***} of S^{***} to the base 100 in Figure 2, with N^{***} satisfying the additional constraint that its grand total is 5,000,000, rather than 4,999,900, thereby also forcing all other marginals in N^{***} to be multiples of 10^6. (The fact that a feasible solution exists to a rounding problem with this additional constraint on the grand total was established by Cox and Ernst (1982).) S in Figure 1 can then be obtained from N^{***} by dividing by 10^6.

The second rounding error problem for this example occurred in the recursive computation of A_k for $k>1$, which can result in rounding error if A_k is computed using real or even double precision arithmetic. To illustrate for the example in Figure 3, the array A_2 with rounding error was obtained from A_1 and N_1 by converting both arrays to double precision, dividing the elements of these arrays by 10,000 and then computing A_2 from (2.4) and (2.6) in double precision arithmetic and rounding to four places.

0.0000	0.0000	0.3579	0.6421	0.0000	1.0000
0.6108	0.3328	0.0000	0.0000	0.0564	1.0000
0.0000	0.4579	0.5421	0.0000	0.0000	1.0000
0.3892	0.2093	0.1001	0.3579	0.9436	2.0001
1.0000	1.0000	1.0001	1.0000	1.0000	5.0001

Figure 3. A_2 with Rounding Error

This rounding error problem can be avoided by performing the recursive computation in a different manner. For $k>1$, N_k is obtained by computing integers, b_k, and integer arrays, $A'_k=(a'_{ijk})$ and $N'_k=(n'_{ijk})$, from $b_{k\text{-}1}$, A'_{k-1}, N'_{k-1}, using integer arithmetic only, as will be described, and then letting

$$N_k=(n'_{ijk}/b_k). \tag{4.1}$$

b_1, A'_1, N'_1 are obtained as described earlier in this section, with b_1=10,000 for this example. For k>1 let

$$b_k=\max\{|n'_{ij(k-1)}-a'_{ij(k-1)}| \;:\; i=1,...,m+1,\; j=1,...,n+1\} \tag{4.2}$$

$$a'_{ijk}=(n'_{ij(k-1)}/b_{(k-1)})b_k+a'_{ij(k-1)}-n'_{ij(k-1)}, \tag{4.3}$$

with only integer arithmetic used in the computation of (4.2) and (4.3). Then let N'_k be a zero-restricted controlled rounding of A'_k to the base b_k.

It can readily be shown that computing the sequence $N_1,...,N_l$ in this manner is equivalent to obtaining this sequence from (2.4) and (2.6), except that using (4.2) and (4.3) avoids rounding error and thus the marginals of $N_1,...,N_l$ computed from (4.2), (4.3) are always the same as the marginals of S. The fact that

$$d_{k\text{-}1}=b_k/b_{k\text{-}1} \tag{4.4}$$

can be used in establishing the equivalence of the two methods of computing $N_1,...,N_l$.

Rounding error also can be avoided in the computation of the sequence of probabilities $p_1,\ldots,p_l$ using the following approach. (2.5) can be rewritten as

$$\begin{aligned} p_k &= (1-d_1) \text{ if } k{=}1 \\ &= \left(\prod_{i=1}^{k-1} d_i\right)(1-d_k) \quad \text{if } k{>}1, \end{aligned} \tag{4.5}$$

where a recursive proof can be used to establish the equivalence of (2.5) and (4.5). Then (4.4), (4.5) yield

$$p_k = \frac{b_k - b_{k+1}}{b_1}, \quad k \geq 1.$$

Thus, for the illustrative example, since b_1=10,000, p_k can be obtained without rounding error by computing the four digit integer $b_k - b_{k+1}$, with possible leading 0's, and placing a decimal point in front of the integer.

Note also that l is the first integer k for which $b_{k+1}=0$.

For the example, l=6, A'_2-A'_6 and N'_2-N'_6 are presented in Figures 4 and 5, respectively, while b_2=5776, b_3=4567, b_4=2500, b_5=578, b_6=252, p_1=.4224, p_2=.1209, p_3=.2067, p_4=.1922, p_5=.0326, p_6=.0252. N_2-N_6 are omitted, but can be obtained by using (4.1).

$$A_2' = \begin{array}{rrrrr|r} 0 & 0 & 2067 & 3709 & 0 & 5776 \\ 3528 & 1922 & 0 & 0 & 326 & 5776 \\ 0 & 2645 & 3131 & 0 & 0 & 5776 \\ 2248 & 1209 & 578 & 2067 & 5450 & 11552 \\ \hline 5776 & 5776 & 5776 & 5776 & 5776 & 28880 \end{array}$$

$$A_3' = \begin{array}{rrrrr|r} 0 & 0 & 2067 & 2500 & 0 & 4567 \\ 2319 & 1922 & 0 & 0 & 326 & 4567 \\ 0 & 2645 & 1922 & 0 & 0 & 4567 \\ 2248 & 0 & 578 & 2067 & 4241 & 9134 \\ \hline 4567 & 4567 & 4567 & 4567 & 4567 & 22835 \end{array}$$

$$A_4' = \begin{array}{rrrrr|r} 0 & 0 & 0 & 2500 & 0 & 2500 \\ 252 & 1922 & 0 & 0 & 326 & 2500 \\ 0 & 578 & 1922 & 0 & 0 & 2500 \\ 2248 & 0 & 578 & 0 & 2174 & 5000 \\ \hline 2500 & 2500 & 2500 & 2500 & 2500 & 12500 \end{array}$$

$$A_5' = \begin{array}{rrrrr|r} 0 & 0 & 0 & 578 & 0 & 578 \\ 252 & 0 & 0 & 0 & 326 & 578 \\ 0 & 578 & 0 & 0 & 0 & 578 \\ 326 & 0 & 578 & 0 & 252 & 1156 \\ \hline 578 & 578 & 578 & 578 & 578 & 2890 \end{array}$$

$$A_6' = \begin{array}{rrrrr|r} 0 & 0 & 0 & 252 & 0 & 252 \\ 252 & 0 & 0 & 0 & 0 & 252 \\ 0 & 252 & 0 & 0 & 0 & 252 \\ 0 & 0 & 252 & 0 & 252 & 504 \\ \hline 252 & 252 & 252 & 252 & 252 & 1260 \end{array}$$

Figure 4. A_2'–A_6'

$$N_2' = \begin{array}{rrrrr|r} 0 & 0 & 0 & 5776 & 0 & 5776 \\ 5776 & 0 & 0 & 0 & 0 & 5776 \\ 0 & 0 & 5776 & 0 & 0 & 5776 \\ 0 & 5776 & 0 & 0 & 5776 & 11552 \\ \hline 5776 & 5776 & 5776 & 5776 & 5776 & 28880 \end{array}$$

$$N_3' = \begin{array}{rrrrr|r} 0 & 0 & 4567 & 0 & 0 & 4567 \\ 4567 & 0 & 0 & 0 & 0 & 4567 \\ 0 & 4567 & 0 & 0 & 0 & 4567 \\ 0 & 0 & 0 & 4567 & 4567 & 9134 \\ \hline 4567 & 4567 & 4567 & 4567 & 4567 & 22835 \end{array}$$

$$N_4' = \begin{array}{rrrrr|r} 0 & 0 & 0 & 2500 & 0 & 2500 \\ 0 & 2500 & 0 & 0 & 0 & 2500 \\ 0 & 0 & 2500 & 0 & 0 & 2500 \\ 2500 & 0 & 0 & 0 & 2500 & 5000 \\ \hline 2500 & 2500 & 2500 & 2500 & 2500 & 12500 \end{array}$$

$$N_5' = \begin{array}{rrrrr|r} 0 & 0 & 0 & 578 & 0 & 578 \\ 0 & 0 & 0 & 0 & 578 & 578 \\ 0 & 578 & 0 & 0 & 0 & 578 \\ 578 & 0 & 578 & 0 & 0 & 1156 \\ \hline 578 & 578 & 578 & 578 & 578 & 2890 \end{array}$$

$$N_6' = \begin{array}{rrrrr|r} 0 & 0 & 0 & 252 & 0 & 252 \\ 252 & 0 & 0 & 0 & 0 & 252 \\ 0 & 252 & 0 & 0 & 0 & 252 \\ 0 & 0 & 252 & 0 & 252 & 504 \\ \hline 252 & 252 & 252 & 252 & 252 & 1260 \end{array}$$

Figure 5. N_2'–N_6'

5. Variances for the Controlled Selection Procedure

In this section, variance formulas are derived for estimators of total for both designs for the sampling procedure detailed in Section 3, under the assumption that a census is conducted in the sample PSUs. If the sample PSUs are subsampled, then these formulas represent the between PSUs component of variance. Let X denote the total value over the entire population for a characteristic of interest and let X_{iju} denote the total for PSU B_{iju} for each $(i,j,u) \in T$. For $\alpha=1,2$, let $\hat{X}_{\alpha}$ denote the usual estimator for X for design α corresponding to probability proportional to size sampling, that is

$$\hat{X}_{\alpha} = \sum \frac{X_{iju}}{P_{iju\alpha}},$$

where the summation is over all (i,j,u) such that $B_{iju} \in G_{\alpha}$. For (i,j,u), $(i^*,j^*,u^*) \in T$, $(i,j,u) \neq (i^*,j^*,u^*)$, $\alpha = 1,2$, let

$$\pi_{ijui^*j^*u^*\alpha} = P(B_{iju}, B_{i^*j^*u^*} \in G_{\alpha}).$$

Finally, for each i,j,i^*,j^* for which $i \le m$, $i^* \le m$, $j \le n$, $j^* \le n$, let $r_{iji^*j^*} = P(n_{ijk} = n_{i^*j^*k} = 1)$. Note that $r_{iji^*j^*}$ is the sum of p_k over all k for which $n_{ijk} = n_{i^*j^*k} = 1$.

Then from Raj (1968, p. 54),

$$V(\hat{X}_\alpha) = \frac{1}{2} \sum_{\substack{(i,j,u),(i^*,j^*,u^*)\in T \\ (i,j,u)\neq(i^*,j^*,u^*)}} (P_{iju\alpha}\, P_{i^*j^*u^*\alpha} - \pi_{ijui^*j^*u^*\alpha}) \left(\frac{X_{iju}}{P_{iju\alpha}} - \frac{X_{i^*j^*u^*}}{P_{i^*j^*u^*\alpha}} \right)^2 . \tag{5.1}$$

Consequently, it is only necessary to show how to compute $\pi_{ijui^*j^*u^*\alpha}$ for each (i,j,u), $(i^*,j^*,u^*)\in T$, $(i,j,u)\neq(i^*,j^*,u^*)$. To do this for α=2, first observe that $\pi_{ijui^*j^*u^*2}=0$ if $j=j^*$. Consequently, it may be assumed from now on that $j\neq j^*$. Then to obtain $\pi_{ijui^*j^*u^*2}$, observe that both B_{iju} and $B_{i^*j^*u^*}$ can be in G_2 if, for some k, either

$$n_{ijk} = n_{i^*j^*k} = 1,\ n_{mjk} = n_{i^*j^*k} = 1,\ n_{ijk} = n_{mj^*k} = 1 \text{ or } n_{mjk} = n_{mj^*k} = 1,$$

which combined with (3.9) and (3.10) yield the four terms in the following expression:

$$\pi_{ijui^*j^*u^*2} = r_{iji^*j^*} \frac{P_{iju3}}{s_{ij}} \frac{P_{i^*j^*u^*3}}{s_{i^*j^*}}$$

$$+ r_{mji^*j^*} \frac{(P_{iju2} - P_{iju3})}{s_{mj}} \frac{P_{i^*j^*u^*3}}{s_{i^*j^*}}$$

$$+ r_{ijmj^*} \frac{P_{iju3}}{s_{ij}} \frac{(P_{i^*j^*u^*2} - P_{i^*j^*u^*3})}{s_{mj^*}}$$

$$+ r_{mjmj^*} \frac{(P_{iju2} - P_{iju3})}{s_{mj}} \frac{(P_{i^*j^*u^*2} - P_{i^*j^*u^*3})}{s_{mj^*}}. \qquad (5.2)$$

The only differences in the expression for $\pi_{ijui^*j^*u1}$, which is obtained similarly, are that the subscripts mj, mj^* and 2 are replaced by the subscripts in, i^*n and 1, respectively, and that $\pi_{ijui^*j^*\mathrm{u}^*1}=0$ if $i=i^*$.

Note, in the special case when $P_{iju1} \leq P_{iju2}$, and hence $P_{iju3} = P_{iju1}$, for all $(i,j,u) \in T$, it follows that the last three terms in the expression for $\pi_{ijui^*j^*u^*1}$ drop out, and therefore,

$$\pi_{ijui^*j^*u^*1} = r_{iji^*j^*} \frac{P_{iju1}}{s_{ij}} \frac{P_{i^*j^*u^*1}}{s_{i^*j^*}}.$$

All four terms in $\pi_{ijui^*j^*u^*2}$ remain, although now 1 can be substituted for 3 throughout (5.2).

Note that (5.2), and hence (5.1), are different for the controlled selection procedure than if independent sampling is used to select the sample PSUs for each design. In the latter case, $\pi_{ijui^*j^*u^*\alpha} = P_{iju\alpha} P_{i^*j^*u^*\alpha}$ if either $\alpha=1$ and $i \neq i^*$, or if $\alpha=2$ and $j \neq j^*$, and hence there is no between strata component of variance for independent sampling.

We next consider further the question of whether controlled selection or independent selection should yield lower variances. Note that for each $(i,j,u) \in T$,

$$\sum_{\substack{(i,{}^*j,{}^*u^*)\in T \\ (i,{}^*j,{}^*u^*)\neq(i,j,u)}} \pi_{ijui^*j^*u^*\alpha} \begin{array}{ll} = (m'-1)\ P_{iju1} & \text{if } \alpha=1, \\ = (n'-1)\ P_{iju2} & \text{if } \alpha=2, \end{array}$$

for both controlled selection and independent selection (see Raj (1968, p.54)), and hence

$$\sum_{\substack{(i,j,u),(i,{}^*j,{}^*u^*)\in T \\ (i,j,u)\neq(i,{}^*j,{}^*u^*)}} (P_{iju\alpha}\ P_{i^*j^*u\alpha} - \pi_{ijui^*j^*u^*\alpha})$$

is the same for both procedures. Consequently, there is no reason to expect the variances for one procedure to be higher or lower than the other unless the relationship between the $P_{iju\alpha}\ P_{i^*j^*u^*\alpha}$ - $\pi_{ijui^*j^*u^*\alpha}$ and the $(X_{iju}/P_{iju\alpha}$ - $X_{i^*j^*u^*}/P_{i^*j^*u^*\alpha})^2$ factors differs for the two procedures. However, one might surmise from the following argument that controlled selection tends to yield lower variances than independent selection for D_2 (and analogously for D_1). For consider (i,j,u), $(i^*,j^*,u^*)\in T$ with $i=i^*$, $j\neq j^*$. Then $\pi_{ijui^*j^*u^*1}$ may tend to be relatively small for such a pair of PSUs since $r_{iji^*j^*}$, and hence the first term in (5.2), are 0. (That is, the probability that a pair of PSUs from the same D_1 stratum, but different D_2 strata, are both D_2 sample PSUs tends to be smaller under controlled selection than independent selection.) Furthermore, assuming the characteristic of interest is well correlated with the D_1 stratification variables, $(X_{iju}/P_{iju\alpha}$ - $X_{i^*j^*u^*}/P_{i^*j^*u^*\alpha})^2$, $\alpha=1,2$, tends to be small for such pairs of PSUs since they are both in the same D_1 stratum. Thus, controlled selection may result in many pairs of PSUs with a large value for $P_{iju1}\ P_{i^*j^*u^*1}$ - $\pi_{ijui^*j^*u^*1}$ and a small value for $(X_{iju}/P_{iju1}$ - $X_{i^*j^*u^*}/P_{i^*j^*u^*1})^2$, a combination which tends to lower variances.

The supposition that controlled selection tends to produce lower variances than independent selection is one of the issues considered in the empirical comparisons presented in the next section.

6. Application to Proposed Expansion of the Current Population Survey

The proposed, but since cancelled, expansion of the CPS would have been an important application of the controlled selection procedure described in the preceding sections. The following is a general outline of this proposal. (For further details see Tupek, Waite and Cahoon (1990).) Beginning in 1994, a redesign of the CPS, based on 1990 census data, is scheduled to be phased in (the D_1 design). The reliability requirements for the redesign are to be approximately the same as for the 1980s design, which has precision requirements for monthly estimates for the nation and the 11 largest states, and for annual estimates for the remaining states and the District of Columbia. Beginning in 1996, if the proposal had been implemented, a sample expansion (the D_2 design) would have taken place to meet reliability requirements for monthly estimates for all 50 states and the District of Columbia.

Four methods have been investigated for selection of the D_2 sample PSUs for this application. In addition to controlled selection, they are the independent sample and independent supplement, both described in Chandhok, Weinstein and Gunlicks (1990), and multiple workloads (Weidman and Ernst 1991). The independent sample method selects the D_2 sample PSUs from an optimal D_2 stratification independently of the D_1 sample PSUs. The independent supplement method includes all D_1 sample PSUs in D_2 and selects additional PSUs for inclusion in the D_2 sample independently from a second supplemental stratification. Mutiple workloads also includes all D_1 sample PSUs and then selects additional PSUs for inclusion in the D_2 sample from the D_1 strata, in a manner that conditions the selections of these additional PSUs on the D_1 sample PSUs.

Among these four procedures, independent selection has the drawback that it will generally result in some D_1 sample PSUs being dropped from the D_2 sample, a feature which undesirably impacts on field operations. Independent selection and independent supplement both might be expected to result in larger variances for D_2, since they do not select D_2 sample PSUs from an optimal D_2 stratification.

The controlled selection approach of this paper with assumption (1.1) can be used as a procedure for simultaneously selecting sample PSUs for both designs while avoiding both of these problems. To use this procedure, first obtain optimal stratifications for D_1 and D_2. Then the controlled selection procedure results in a set of sample PSUs for D_1 and D_2 satisfying (1.2) and (1.3).

An empirical investigation was undertaken to compare variances using the four approaches to selection of PSUs for the proposed CPS expansion. For this comparison of the variances, the D_1 and D_2 stratifications were obtained using several labor force characteristics from the 1980 census as stratification variables. 1980 census data was used since 1990 data was not available at the time the stratification was done. A modified Friedman-Rubin clustering algorithm (Kostanich et al. 1981) was used to obtain the stratifications. The D_1 and D_2 stratifications and the controlled selection were performed separately for each state.

The variables used here to compare the variances of the four procedures are number of unemployed persons and number of persons in the civilian labor force. The comparisons were done only for the 31 states listed in Table 1. Of the remaining 20 states (counting the District of Columbia), the 11 largest were omitted since the precision requirements for this study, and hence the stratifications, were the same for D_1 and D_2. Eight states were omitted because they consisted entirely of self-representing PSUs for D_1. For these 19 states, variances for all four procedures would be identical for both D_1 and D_2. Finally, Alaska was omitted because of problems with the data files.

For each state and each characteristic, variances were computed for each of the two designs and each of the four selection procedures using 1970 census data, which was chosen to simulate a 10-year lag between the data used in the stratifications and the collection of the survey data, which is roughly the anticipated average lag time for the D_1 and D_2 designs. The total number of sample persons assumed for the D_2 design for each of the four methods is the number needed by the independent sample method to meet the proposed D_2 reliability requirements. This resulted in the same estimate of within PSUs variances for each of the four methods, which was obtained by multiplying the simple random sample variance by a fixed design effect, thus allowing the variance comparisons to be made on the basis of between PSUs variances only.

Table 1 can be used to compare the between PSU variances for the four procedures. Each entry in numerical columns 1, 2, 5 and 6 is a ratio of the between PSUs variance for the controlled selection procedure to the between PSUs variance for independent sample for the indicated state, characteristic and design. Each entry in the remaining four columns is the ratio of the between PSUs variance of either multiple workloads or independent supplement to independent sample for the D_2 design. (The D_1 variances are the same for all the methods except controlled selection since the D_1 sample is selected in the same way for the other three methods.) Each entry in the last row of a column is the arithmetic mean of the entries in the preceding rows of that column.

It can be observed from Table 1 that the D_2 between PSUs variances are generally considerably lower for controlled selection than for either multiple workloads or independent supplement. Furthermore, the means of the ratios in the first four columns are all relatively close to 1. These numbers do not provide support for the supposition in Section 5 that controlled selection tends to produce lower variances than independent selection, but instead are more in line with the hypothesis that neither method is superior to the other in terms of between PSUs variances.

An additional observation is that the deviations from 1 of the ratios of the between PSUs variances for controlled selection to independent selection are generally smaller on a state-by-state basis for D_2 than D_1, an observation for which these are at least two explanations. First, for the controlled selection procedure with assumption (1.1), no two PSUs in the same D_2 stratum can be in G_1. However, there can be as many PSUs in G_2 from a single D_1 stratum as there are D_2 strata with PSUs from that D_1 stratum, provided this does not violate the requirement imposed by (3.6) on the maximum number of PSUs in $G_2 \sim G_1$. Thus, the restrictions imposed by the controlled selection procedure on the possible sets of sample PSUs are more restrictive for D_1 than for D_2, which partially explains the smaller deviations of the ratios for D_2.

The second reason for the smaller deviations for D_2 is that many of the D_2 strata in this application consisted entirely of PSUs from a single D_1 stratum. If the j-th D_2 stratum is such a stratum, then for this j, $\pi_{ijui^*j^*u^*2} = P_{iju2} \, P_{i^*j^*u^*2}$ if $j^* \neq j$ and $\pi_{ijui^*j^*u^*2} = 0$ if $j^* = j$, for all distinct triples (i,j,u), $(i^*,j^*,u^*) \in T$ for both controlled selection and independent selection, and thus the contribution to (5.1) from all such pairs is the same for both of these procedures for D_2. No analogous relationship holds for D_1.

Although the between PSUs variances for controlled selection are considerably lower than those for multiple workloads and independent supplement, the same is not true for total variances, since within PSUs variance, which is estimated to be the same for all the procedures, is the dominant component of total variance for each of these estimators. For example, Weidman and Ernst (1991) present the analogous table to Table 1 for total variance. None of the entries in the bottom row of that table exceed 1.130.

In summary, based on this limited study, if the proposed CPS expansion had taken place and it had been required that the D_1 sample PSUs be a subset of the D_2 sample PSUs, then among the three methods considered here which satisfy this requirement, controlled selection is the clear choice if minimization of between PSUs variances is the chief criterion. However, as previously mentioned, controlled selection is not usable in applications where the D_2 sample PSUs are selected subsequent to selection of the D_1 sample PSUs. Furthermore, unless one is willing to ignore the effect on the variances of the between strata variance component induced by controlled selection, variance estimation would be more complex than for some other approaches to selection of PSUs.

Acknowledgement

The author would like to thank Todd Williams who did the programming that produced all the tables and figures that appear in this paper.

7. References

Causey, B.D., Cox, L.H., and Ernst, L.R. (1985). Applications of Transportation Theory to Statistical Problems. Journal of the American Statistical Association, 80, 903-909.

Chandhok, P., Weinstein, R., and Gunlicks, C. (1990). Augmenting a Sample to Satisfy Subpopulation Requirements. Proceedings of the Section on Survey Research Methods, American Statistical Association, 696-701.

Cox, L.H., and Ernst, L.R. (1982). Controlled Rounding. INFOR, 20, 423-432.

Ernst, L.R. (1986). Maximizing the Overlap Between Surveys When Information Is Incomplete. European Journal of Operational Research, 27, 192-200.

Kostanich, D., Judkins, D., Singh, R., and Schautz, M. (1981). Modification of Friedman-Rubin's Clustering Algorithm for Use in Stratified PPS Sampling. Proceedings of the Section on Survey Research Methods, American Statistical Association, 285-290.

Pruhs, K. (1989). The Computional Complexity of Some Survey Overlap Problems. Proceedings of the Section on Survey Research Methods, American Statistical Association, 747-752.

Raj, D. (1968). Sampling Theory. New York: McGraw Hill.

Tupek, A.R., Waite P.J., and Cahoon, L.S. (1990). Sample Expansion Plans for the Current Population Survey. Proceedings of the Section on Survey Research Methods, American Statistical Association, 72-77.

Weidman, L., and Ernst L.R. (1991). Multiple Workloads per Stratum Sampling Designs. Proceedings of the Section on Survey Research Methods, American Statistical Association, 443-448.

Table 1

Ratios of Between PSU Variances for Other Options to the Independent Sample

	Unemployed				Civilian Labor Force			
	D_1	D_2			D_1	D_2		
State	CS	CS	IS	MW	CS	CS	IS	MW
Alabama	0.71	1.12	3.36	2.97	1.24	1.03	2.87	5.33
Arizona	0.83	1.07	4.21	18.63	0.61	0.89	4.68	1.84
Arkansas	0.88	1.06	2.67	0.53	1.44	0.99	1.57	0.33
Colorado	1.60	1.01	4.32	3.59	0.78	1.00	5.31	2.33
Georgia	0.80	1.04	2.42	1.25	1.20	1.02	1.91	0.64
Idaho	0.67	1.10	13.47	2.96	0.88	1.06	4.64	2.05
Indiana	0.57	0.89	4.00	4.52	1.34	1.47	3.90	1.08
Iowa	1.23	0.78	3.63	1.10	1.60	0.83	4.29	1.09
Kansas	0.73	0.95	3.49	1.89	0.99	0.92	1.71	0.85
Kentucky	2.08	1.08	3.35	4.21	0.42	0.84	3.67	3.13
Louisiana	2.27	0.95	2.65	3.88	0.76	0.97	2.63	2.10
Maryland	0.91	1.00	4.62	1.09	0.85	1.00	4.92	0.04
Minnesota	0.88	1.12	1.51	1.69	0.82	0.87	2.75	1.39
Mississippi	1.12	0.93	4.84	1.28	1.66	1.03	3.70	0.59
Missouri	0.56	0.95	3.99	2.17	0.55	1.09	3.25	0.61
Montana	0.82	0.88	2.16	2.53	1.24	1.19	6.24	0.58
Nebraska	0.95	0.99	2.83	1.52	0.78	1.00	2.78	0.62
Nevada	1.12	1.02	5.81	0.78	0.66	0.86	15.02	0.35
New Mexico	0.75	0.91	3.06	7.27	0.63	1.41	4.13	2.98
North Dakota	0.71	0.91	6.34	1.85	0.88	0.72	3.09	0.57
Oklahoma	1.14	1.02	2.72	2.27	0.43	0.83	3.86	0.88
Oregon	0.74	0.89	2.25	2.72	0.63	0.98	5.40	0.61
South Carolina	1.25	1.17	3.64	0.68	0.83	1.10	9.67	1.56
South Dakota	0.85	0.90	2.53	1.56	0.99	0.95	2.50	0.63
Tennessee	1.11	1.10	8.16	1.77	0.58	1.00	3.01	0.76
Utah	0.93	0.97	2.66	1.19	1.48	0.94	2.67	0.73
Virginia	1.11	0.95	2.87	0.87	2.10	1.33	3.07	0.44
Washington	1.14	0.94	3.29	1.78	0.54	1.25	5.86	1.48
West Virginia	3.05	0.93	1.60	4.23	2.13	1.12	0.37	1.22
Wisconsin	1.60	0.96	2.17	1.62	0.92	0.98	2.94	0.75
Wyoming	0.37	0.67	5.69	5.04	0.43	1.03	35.91	12.08
Mean	1.08	0.98	3.88	2.89	0.98	1.02	5.11	1.60

CS = Controlled Selection
IS = Independent Supplement
MW = Multiple Workloads