A System for Replacing Primary Sampling Units When the Units Have Been Exhausted

Charles H. Alexander, Lawrence R. Ernst and Michael E. Haas *
Bureau of the Census

Abstract

This paper describes a theoretical investigation of alternative methods for replacing primary sampling units (PSU's) when the units are not large enough to last for the lifetime of a continuing survey. The aim is to keep the sample unbiased while minimizing the expected number of replacements which are necessary. The present method used for the Census Bureau's recurring surveys is compared to a general class of alternatives. It is shown that when the present method is applied using optimal groupings of PSU's, it is at least as good as any of this class of alternatives.

1. INTRODUCTION

For the Current Population Survey (CPS) and other Census Bureau current surveys after the 1980's sample redesign, primary sampling units (PSU's) be counties or groups of counties. The most populous PSU's will be in sample with probability one (self-representing or SR PSU's). Non-self-representing (NSR) PSU's will be grouped into strata of similar PSU's and one or two PSU's will be selected from each stratum with probability proportional to size. Regardless of which PSU's are selected for the sample, approximately the same number of housing units will be sampled; this leads to an equal-probability or "self-weighting" sample. In most NSR strata, a single interviewer will handle the entire selected PSU. The stratum sizes and sampling rates are selected so as to provide an efficient workload for the interviewer while yielding the desired national sample size.

Because of these features of the current surveys, it is sometimes necessary to replace small sample PSU's during the life of the survey. This happens when the assigned sampling rate is so large that there are not enough housing units in the county to supply the necessary sample for the decade, or until the next sample selection. To avoid undue respondent burden, one should avoid including the same address in sample twice during the decade. The assigned sampling rate cannot be changed without losing the self-weighting property of the design, so replacement is the only course left. The problem is how to make such replacements without biasing the sample.

The 1970's CPS design dealt with this problem through the use of "rotation clusters." Small PSU's were formed into groups so that each group was large enough to supply housing units for the life of the sample. If any PSU in the group was selected, then interviewing rotated among the PSU's in the cluster during the decade. The length of the time each PSU spent in sample was chosen so that the PSU's in sample at any future time constitute an unbiased sample of PSU's.

This rotation of PSU's is expensive and inconvenient, usually requiring the training of a new interviewer. For this reason, it is desirable to reduce the expected number of rotations, while keeping the sample unbiased.

As part of the current surveys redesign, an attempt was made to develop an alternative unbiased method which reduced the expected number of rotations. A class of unbiased methods was developed, which we call the Length of Time in Sample Methods (LTSM). Depending on how the rotation clusters are formed, these methods may reduce the expected number of rotations compared to the present method. However, when the rotation clusters are formed in an optimal fashion, the present method always does at least as well as any LTSM.

For this paper, we have developed an idealized version of the present method, which we call the Random Arc Method (RAM). The present method has not been documented in its entirety. The resulting rotation schedules are given in Brooks(1972). Related work is contained in Brooks (1971) and Brooks and Hanson (1975). The principle themes of this paper are to determine an optimal clustering for RAM, i.e., one which minimizes the expected number of rotations when RAM is used, and to establish that with an optimal clustering RAM reduces the expected number of rotations compared to any LTSM. The 1970's sample selection did not use the optimal clustering in all cases.

While RAM does not always minimize the expected number of rotations among all methods which keep the sample unbiased, it is shown that it gives at most one more than the minimum possible expected number. RAM is known not to be optimal for certain distributions of PSU population (see example 3.1); however so far there does not appear to be a generally applicable method which is superior.

The LTSM are presented to document our investigation and because if a different cost function is used they may be admissible competitors to RAM. For example, if rotation of a PSU into sample for the second time is regarded as cost-free because a new interviewer need not be trained, LTSM is in some cases superior to RAM (see example 3.3). However, for the Census Bureau's current surveys, the appropriate cost function is based on the expected number of PSU's.

In Section 2 we state the model and briefly describe the steps involved in the clustering and rotation of PSU's. In Section 3 we define the methods under consideration and present several theorems on the conditional expected number of rotations for the chosen cluster. In Section 4 we prove theorems on the expected number of rotations for the entire stratum, and obtain the two results described above. Finally, in Section 5, some extensions of the method are discussed.

Although a final decision has not as yet been made on whether a one or two PSU's per stratum design will be employed, for simplicity a one PSU per stratum design will be assumed throughout the remainder of the first four sections of this paper. In Section 5 the two PSU's per stratum case will be briefly discussed.

2. THE MODEL AND OUTLINE OF BASIC STEPS

For the selection of a sample for a single survey the problem may be considered separately for each stratum S of PSU's.

Assume that the survey of interest interviews housing units continuously and uniformly in the interval $[0,T)$ months, such that for $0 < t \leq T$, nt/T units are interviewed in $[0,t)$. Thus in the entire life of the sample the number of units interviewed is n. This model is only an approximation; for example, for

some surveys interviewing occurs only at the start of a month, while for others rotation is feasible only at certain intervals.

There are three steps in the process of rotation of PSU's for all procedures under consideration. First the set of PSU's in the stratum is partitioned into $\ell \geq 1$ clusters, $C_1,\ldots,C_\ell$, each consisting of one or more PSU's. The total number of housing units in each cluster must , of course, be at least n. The formation of clusters will be discussed in Section 4.

The second step is to select one of these clusters PPS.

The final step is to specify for any time $t \in [0,T)$ which PSU in the chosen cluster is in sample at that time. Procedures for such designation will be discussed next in Section 3.

Note that even though it is possible to devise procedures which do not involve the clustering of PSU's, such procedures may be viewed as a special case of the three step process just described, for which there is only one cluster consisting of all PSU's in the stratum.

The aim of all the methods is to make each PSU's expected share of the T interview months equal to its share of the cluster population. For LTSM, the PSUs' lengths of time in sample are chosen randomly subject to certain constraints. Some PSU's in the selected cluster may have a positive probability of spending no time in sample. To make up for this, the PSU's share of the T months will sometimes be greater than its share of the cluster population. Once the lengths of time have been selected, the order in which the PSU's enter the sample is determined. The first PSU in the sample spends part of its time at the beginning of the decade and part at the end; the others are in sample for a single interval of time. LTSM are defined in more detail in Section 3.2.

RAM simultaneously determines the order of entry of the PSU's into sample and the length of time in sample, using a simple geometric technique which is described in Section 3.3.

3. ROTATION OF PSU's IN SELECTED CLUSTER

We fix notation in 3.1, define LTSM and RAM in 3.2 and 3.3 respectively, present the main results of this section in 3.4, and discuss the methods and results in 3.5.

3.1 NOTATION AND TERMINOLOGY

Let C denote the selected cluster, consisting of PSU's $A_1,\ldots,A_k$. Let the population of A_i be $\pi(A_i)$ and let

$$\pi(C) = \sum_{i=1}^{k} \pi(A_i) .$$

A_i is said to be a small PSU if $\pi(A_i)<n$, a large PSU otherwise. Let m be the number of small PSU's in C and assume that $A_1,\ldots,A_m$ are the small PSU's. It will be understood that throughout this section all probabilities and expected values are conditional on C being the sample cluster. They will be denoted P_c and E_c respectively.

A **rotation schedule** ω for C specifies for $t \in [0,T)$ which PSU is in sample at time t. Let I_t be the PSU which is in sample at time t. It is assumed for every rotation schedule and each i=1,...k, that $\{t: I_t = A_i\}$ is either empty or a finite union of nonempty intervals; the sum of the lengths of these intervals, i.e. the total amount of time during which A_i is in sample, will be denoted by $\tau(A_i)$. A **rotation method** is a method of randomly selecting a rotation schedule. A rotation method is defined to be **unbiased** if

$$P_c(I_t = A_i) = \frac{\pi(A_i)}{\pi(C)} ,$$

$$i=1,\ldots,k, \quad t \in [0,T) . \qquad (3.1)$$

Let $R(\omega)$ denote the number of times PSU's are rotated into sample during the entire life of the sample using rotation schedule ω. (When the first PSU enters at time t=0 this is not counted as a rotation.) Finally, throughout this section we assume that $k \geq 2$, since otherwise there is only one possible rotation schedule ω, for which $R(\omega)=0$.

3.2 LENGTH OF TIME IN SAMPLE METHODS (LTSM)

This class of methods consists of the following. For each PSU A_i, assign a number $\tau(A_i) \in [0,T]$ giving the length of time in sample for PSU A_i, (hence the name for this class of methods). The values $\tau(A_1),\ldots,\tau(A_k)$ are random variables from some specified joint distribution. It is required that

$$\sum_{i=1}^{k} \tau(A_i) = T, \qquad (3.2)$$

and also that

$$\tau(A_i) \leq \frac{\pi(A_i)T}{n}. \qquad (3.3)$$

(The latter condition is necessary in order that the $n\tau(A_i)/T$ housing units which are interviewed during $\tau(A_i)$ months do not exceed $\pi(A_i)$). Note that (3.2) and (3.3) must be satisfied by all rotation methods, not only LTSM.

If $\tau(A_i) = T$ for some i, then A_i is in sample for the entire time interval $[0,T)$.

Otherwise, once the values $\tau(A_1),\ldots,\tau(A_k)$ have been specified, the s PSU's for which $\tau(A_i) > 0$ are placed in random order. Let $A_{(1)},\ldots,A_{(s)}$ denote the PSU's in the chosen order. Then choose x at random in the interval $[0,T)$. Let u be the smallest positive integer for which

$x < \sum_{j=1}^{u} \tau(A_{(j)})$. Then $A_{(u)}$ will be in sample at the beginning and again at the end of the period, during the intervals $[0, \sum_{j=1}^{u} \tau(A_{(j)})-x)$ and $[T+\sum_{j=1}^{u-1} \tau(A_{(j)})-x, T)$. (If $x = \sum_{j=1}^{u-1} \tau(A_{(j)})$, then $A_{(u)}$ is only in sample once, for the interval $[0,\tau(A_{(u)}))$. However, this event occurs with probability 0, and will be ignored in the subsequent discussion.) The other PSU's A_i for which $\tau(A_i) > 0$ will rotate into sample in the order $A_{(u+1)}, A_{(u+2)},\ldots,A_{(s)}, A_1, A_2,\ldots,A_{(u-1)}$. In general $A_{(i)}$ is in sample for the interval

$$\left[\sum_{j=1}^{i-1}\tau(A_{(j)})-x,\ \sum_{j=1}^{i}\tau(A_{(j)})-x\right),$$

$$i = u+1,\ldots,s, \qquad (3.4)$$

$$\left[\sum_{j=1}^{i-1}\tau(A_{(j)}) + T-x,\ \sum_{j=1}^{i}\tau(A_{(j)}) + T-x\right),$$

$$i = 1,\ldots,u-1. \qquad (3.5)$$

Geometrically, this definition may be viewed as follows. Partition the interval $[0,T)$ into s half-closed, half-open intervals of length $\tau(A_{(1)}),\ldots,\tau(A_{(s)})$, and then bend the interval $[0,T)$ to form a circle of circumference T, with clockwise orientation, partitioned into s arcs. For any $a,b \in [0,T)$ the arc ab will denote the half-closed, half-open arc with initial point a and final point b when viewed clockwise. A starting point x on the circumference is selected randomly according to a uniform distribution. At any time t our location on the circle will be t units measured clockwise from x. If this point is in the i-th arc of the partition then $A_{(i)}$ is in sample at time t. Intervals

$$\left[0,\ \sum_{j=1}^{u}\tau(A_{(j)})-x\right) \text{ and } (3.4) \text{ correspond to}$$

time periods when located in the arc x0, while

$$\text{intervals } (3.5) \text{ and } \left[\sum_{j=1}^{u-1}A_{(j)}+T-x,T\right)$$

correspond to time periods located in the arc 0x. Since the circle is traversed exactly once in the time interval $[0,T)$, at the end of the time interval we are located in the same arc of the partition as in the beginning.

It is clear that for LTSM we have

$$P_c[I_t = A_i \mid \tau(A_i)] = \frac{\tau(A_i)}{T},$$

$$i = 1,\ldots,k, \quad t \in [0,T). \qquad (3.6)$$

Consequently, for such a rotation method to be unbiased it is necessary and sufficient for

$$E_c[\tau(A_i)] = \frac{T\pi(A_i)}{\pi(C)},$$

$$i = 1,\ldots,k, \qquad (3.7)$$

since (3.6) and (3.7) imply (3.1). (Note that (3.7) must hold for all unbiased rotation methods by (3.1).)

In summary, for the class of rotation methods considered, it is only necessary to specify the joint distribution of $\tau(A_1),\ldots,\tau(A_k)$ satisfying (3.2) and (3.3). Provided (3.7) is also satisfied, then the rotation method is unbiased.

3.3 THE RANDOM ARC METHOD (RAM)

To define this method we first let $\lambda(A_i) = T\pi(A_i)/n$ for $i=1,\ldots,k$, and $\lambda(C) = T\pi(C)/n$. I.e., $\lambda(A_i)$ is the time it would take to exahust PSU A_i and $\lambda(C)$ is the time which would be required to exhaust every PSU in the cluster. Note that $\lambda(C) \geq T$.

Place the k PSU's in random order. Let $A_{(1)},\ldots,A_{(k)}$ denote the PSU's in the chosen order. Then choose x at random in the interval $[0,\lambda(C))$. Let u be the smallest positive integer satisfying $x < \sum_{j=1}^{u}\lambda(A_{(j)})$. Then $A_{(i)}$ is defined to be in sample for the following time intervals, some of which may be empty.

$$\left[\sum_{j=1}^{i-1}\lambda(A_{(j)})-x,\ \sum_{j=1}^{i}\lambda(A_{(j)})-x\right) \cap [0,T),$$

$$i = u,\ldots,k,$$

$$\left[\sum_{j=1}^{i-1}\lambda(A_{(j)})+\lambda(C)-x,\ \sum_{j=1}^{i}\lambda(A_{(j)})+\lambda(C)-x\right)$$

$$\cap [0,T), \quad i = 1,\ldots,u.$$

RAM may also be viewed geometrically in terms of a circle, but one of circumference $\lambda(C)$ partitioned into arcs of length $\lambda(A_{(1)}),\ldots,\lambda(A_{(k)})$. A point x is chosen at random along the circumference. The PSU's will come into sample in the order $A_{(u)}, A_{(u+1)},\ldots,A_{(k)},A_{(1)},\ldots,A_{(u)}$. However, when $\lambda(C)>T$, the entire circle will not be traversed, just a random arc of length T (hence the name of the method), and PSU's at the latter part of the order may not actually come into sample. On the other hand, in some cases $A_{(u)}$ and only $A_{(u)}$ may come into sample twice. This is possible only if

$$\sum_{j\neq u}\lambda(A_{(j)}) < T \quad \text{or, equivalently, if}$$

$$\sum_{j\neq u}\pi(A_{(j)}) < n.$$

We next show that RAM is an unbiased rotation method. Clearly (3.2) is satisfied. Furthermore, (3.3) is satisfied since the total amount of time $A_{(j)}$ is in sample does not exceed $\lambda(A_{(j)}) = T\pi(A_{(j)})/n$. Finally, the method is unbiased since for any $t \in [0,T)$, $A_{(i)}$ is in sample at time t if and only if x is in the arc of length $\lambda(A_{(j)})$ obtained by translating the i-th arc of the partition t units counterclockwise. Since the i-th arc has length $\lambda(A_{(i)})$, this event occurs with probability $\lambda(A_{(i)})/\lambda(C) = \pi(A_{(i)})/\pi(C)$, and hence (3.1) follows.

3.4 MAIN RESULTS

Our main results on the expected number of rotations of PSU's for the selected cluster are contained in the following three theorems:

Theorem 3.1. For any unbiased rotation method,

$$E_c[R(\omega)] \geq \frac{nk}{\pi(C)} - 1.$$

Theorem 3.2. For any LTSM,

$$(a)\quad E_c[R(\omega)] \geq \frac{nm}{\pi(C)},$$

and furthermore,

$$(b)\quad E_c[R(\omega)] \geq \frac{n(m+1)}{\pi(C)} \quad \text{if } 0 < \sum_{i=1}^{m} \pi(A_i) < n.$$

Theorem 3.3. For RAM,

$$E_c[R(\omega)] = \frac{nk}{\pi(C)}.$$

Proofs of these three theorems are available from the authors.

3.5 DISCUSSION

By Theorems 3.2 and 3.3, the expected number of rotations for RAM never exceeds that for any LTSM if either $k = m$ (i.e. all PSU's in C are small), or $k = m+1$ and $\sum_{i=1}^{m} \pi(A_i) < n$ (i.e. C contains exactly one large PSU and the total number of housing units in the small PSU's is less than n). Although for some other types of clusters, $E_c[R(\omega)]$ could be less for some LTSM than for RAM, it will be proven in Section 4 that if the optimal clustering is used for RAM, then the expected number of rotations over the entire stratum is less than or equal to the expected number for any LTSM.

From Theorems 3.1 and 3.3 we have that for RAM, $E_c[R(\omega)]$ exceeds the optimal by at most 1. The following two examples illustrate that the lower bound given in Theorem 3.1 can sometimes be attained, but not always.

Example 3.1. Let $k = 2$, $\pi(A_1) = \pi(A_2) = n/2$. Then consider the rotation method such that with probability 1/2, $I_t = A_1$ for $t \in [0,T/2)$ and $I_t = A_2$ for $t \in [T/2,T)$, and with probability 1/2, $I_t = A_2$ for $t \in [0,T/2)$ and $I_t = A_1$ for $t \in [T/2,T)$. Clearly this defines an unbiased rotation method for this cluster with $E_c[R(\omega)] = 1$, which is the lower bound given in Theorem 3.1 for this example. (Note that for RAM, $E_c[R(\omega)] = 2$.)

Example 3.2. Let $k = 2$, $\pi(A_1) = 2n/3$, $\pi(A_2) = n/3$. The lower bound given in Theorem 3.1 is again 1, but it will be shown that for this example $E_c[R(\omega)] > 1$ for all unbiased rotation methods. In fact, there are only four rotation schedules for which $R(\omega) = 1$, namely the schedule $I_t = A_1$ for $t \in [0,2T/3)$ and $I_t = A_2$ for $t \in [2T/3,T)$, the schedule $I_t = A_2$ for $t \in [0,T/3)$ and $I_t = A_1$ for $t \in [T/3,T)$, and the two equivalent schedules in which the first interval is closed and the second is open. However $I_t \neq A_2$ for $t \in (T/3,2T/3)$ for any of these rotation schedules, and hence any unbiased rotation method must assign positive probability to some other rotation schedules, for which $R(\omega) \geq 2$.

RAM gives good results in part because it entirely uses up those PSU's which come into sample, except for the first and last ones. It is desirable to keep PSU's in sample as long as possible when they do come into sample, because then the probability of not coming into sample can be larger, while still satisfying (3.7). In this respect, it is similar to those LTSM which are designed to keep sample PSU's in sample as long as possible. Both LTSM and RAM define a length of time in sample for each PSU and determine a starting point in the initial PSU. The methods differ in that for LTSM the selection of the starting point is made after the lengths of time have been selected, while for RAM the selections are made simultaneously.

As mentioned in the introduction, it may be of interest to ignore rotation into sample for the second time. This is equivalent to minimizing the expected number of PSU's in sample, rather than the expected number of rotations. The following example shows that in some cases the expected number of PSU's in sample may be smaller using LTSM than using RAM.

Example 3.3. Let $k=2$, $\pi(A_1)=n/3$, $\pi(A_2)=5n/3$. In this case $\lambda(A_1)=T/3$, $\lambda(A_2)=5T/3$, and $\lambda(C)=2T$. Using RAM, the expected number of PSU's in sample is 5/3. However, consider the LTSM which gives probability 1/2 to the event $\{\tau(A_1)=0, \tau(A_2)=T\}$ and probability 1/2 to the event $\{\tau(A_1)=T/3, \tau(A_2)=2T/3\}$. For this LTSM, the expected number of PSU's in sample is 3/2. For both methods, $E_c[R(\omega)]=1$.

For both RAM and LTSM, unbiased sample selection is achieved even if the PSU's are not placed in random order. There may be cost savings if replacements are made in a particular order, for example, if geographically contiguous PSU's are placed consecutively around the circle. However, there is a danger that a nonrandom order may introduce undesirable patterns into estimates of change from one time period to another, especially if the order is correlated with the characteristics being measured by the survey.

4. CLUSTERING OF PSU'S

In this section we prove results on the expected number of rotations over the entire stratum S, obtain the optimal clustering for RAM, and make comparisons with LTSM.

We first introduce notation. For $i = 1,\ldots,\ell$ let E_{c_i}, P_{c_i} denote respectively expectation and probability conditioned on C_i being the sample cluster. For any rotation schedule ω, denote by $R(\omega)$ the number of rotations for the entire stratum, which is the same as the number of rotations in the selected cluster.

Note that

$$E[R(\omega)] = \sum_{i=1}^{\ell} \frac{\pi(C_i)}{\pi(S)} E_{c_i}[R(\omega)]. \qquad (4.1)$$

For $i = 1,\ldots,\ell$ let k_i, m_i be respectively the number of PSU's and the number of small PSU's in C_i, and then let $G = \{i: k_i > 1\}$. Finally, let $M = \sum_{i=1}^{\ell} m_i$, $K = \sum_{i \in G} k_i$. Thus M is the number of small PSU's in S, and K is the number of PSU's in S that are in clusters containing more than 1 PSU. Note that since $E_{c_i}[R(\omega)] = 0$ if $k_i = 1$, (4.1) reduces to

$$E[R(\omega)] = \sum_{i \in G} \frac{\pi(C_i)}{\pi(S)} E_{C_i}[R(\omega)]. \quad (4.2)$$

Note also that if $k_i = 1$ then $m_i = 0$, and hence

$$\sum_{i \in G} m_i = M. \quad (4.3)$$

We next prove three theorems concerning $E[R(\omega)]$.

Theorem 4.1. For any unbiased rotation method,

$$E[R(\omega)] \geq \frac{nK - \sum_{i \in G} \pi(C_i)}{\pi(S)}.$$

Proof. This follows from (4.2) and Theorem 3.1, since

$$E[R(\omega)] = \sum_{i \in G} \frac{\pi(C_i)}{\pi(S)} E_{C_i}[R(\omega)] \geq$$

$$\frac{n \sum_{i \in G} k_i}{\pi(S)} - \frac{\sum_{i \in G} \pi(C_i)}{\pi(S)} = \frac{nK - \sum_{i \in G} \pi(C_i)}{\pi(S)}.$$

Theorem 4.2. For any LTSM,

(a) $$E[R(\omega)] \geq \frac{nM}{\pi(S)},$$

and furthermore,

(b) if there exists some C_j for which $m_j \geq 1$ but the total number of housing units in small PSU's of C_j is less than n, then

$$E[R(\omega)] \geq \frac{n(M+1)}{\pi(S)}.$$

Proof. Combine (4.2), Theorem 3.2 (a), and (4.3) to immediately obtain (a).

To prove (b), observe that by (4.2), Theorem 3.2 (a) and (b), and (4.3),

$$E[R(\omega)] = \frac{\pi(C_j)}{\pi(S)} E_{C_j}[R(\omega)] +$$

$$\sum_{\substack{i \in G \\ i \neq j}} \frac{\pi(C_i)}{\pi(S)} E_{C_i}[R(\omega)]$$

$$\geq \frac{n(m_j+1)}{\pi(S)} + \frac{n \sum_{\substack{i \in G \\ i \neq j}} m_i}{\pi(S)} = \frac{n(M+1)}{\pi(S)}.$$

Theorem 4.3. For RAM,

$$E[R(\omega)] = \frac{n \sum_{i \in G} k_i}{\pi(S)} = \frac{nK}{\pi(S)}.$$

Proof. Combine (4.2) and Theorem 3.3.

The following theorem concerning the optimal clustering of PSU's for RAM now follows from Theorem 4.3.

Theorem 4.4. For RAM, a clustering that minimizes $E[R(\omega)]$, and the minimum value of $E[R(\omega)]$ are as follows:

(a) If the total number of housing units in all small PSU's is at least n, then place all the small PSU's in one cluster, and all the large PSU's in clusters consisting of a single PSU. In this case,

$$E[R(\omega)] = \frac{nM}{\pi(S)}.$$

(b) If the total number of housing units in small PSU's is less than n, then place all the small PSU's in one cluster together with one large PSU. Place all the other PSU's in clusters consisting of a single PSU. In this case,

$$E[R(\omega)] = \frac{n(M+1)}{\pi(S)}.$$

Proof. From Theorem 4.3 it follows that minimizing $E[R(\omega)]$ for RAM is equivalent to minimizing K, the number of PSU's in clusters of more than 1 PSU. It is clear that the clusterings described in (a) and (b) do accomplish this. Furthermore, since $K = M$ for the clustering of (a) and $K = M+1$ for the clustering of (b), the given values for $E[R(\omega)]$ in the two cases follow immediately from Theorem 4.3.

Finally, we have the following result concerning the superiority of RAM to any LTSM.

Theorem 4.5. If RAM is used together with the clustering described in Theorem 4.5, then the resulting value of $E[R(\omega)]$ will be no more than for any LTSM, irrespective of the clustering used.

Proof. If S satisfies the conditions of Theorem 4.4 (a) then the result follows from Theorems 4.2 (a) and 4.4 (a). Otherwise, it follows from Theorems 4.2 (b) and 4.4 (b).

5. EXTENSIONS

For application to the sample selection for the Census Bureau's current surveys, some modifications must be made to the methods described in the previous sections. These will be discussed only briefly, for reasons of space.

One modification is necessary because several surveys will be in the field at the same time. If several surveys share a comon stratum, RAM is easily modified by forming clusters large enough to supply sample for all the surveys and then selecting a random arc for each survey. Start each one where the previous one leaves off. This way two surveys are never in the same PSU at the same time. For some surveys, it is preferable to have several surveys in the same PSU, so the same interviewer can be used. The method can be modified to achieve this. If different surveys have different strata, this method can still be applied in some circumstances.

Additional modifications are necessary if for some surveys two PSU's are to be selected from each stratum. This must be done so as to give unbiased sample selection and also to give each pair of PSU's a

positive joint probability so that unbiased variance estimation is possible. This places certain constraints on the sizes of the clusters which may be formed. A modified version of RAM for selecting two PSU's per stratum has been developed and will be documented in an internal Census Bureau memorandum.

REFERENCES

Brooks, Camilla (1971), "The Unbiased Rotation of MLS and CPS Hits Between Sample PSU's," U.S. Bureau of the Census memorandum, dated February 24, 1971, for the record.

Brooks, Camilla (1972), "Sample PSU's in the 461 Area Redesign of the Current Population Survey," U.S. Bureau of the Census memorandum, dated March 29, 1972, to Gary Shapiro.

Brooks, Camilla and Hanson, Robert H. (1975), "Rotation of PSU's for A and C Design Samples," U. S. Bureau of the Census memorandum, dated March 26, 1975, for the record.

---

* The authors would like to thank David Judkins for his valuable assistance in the early stage of this research.