

# OPTIMAL ALLOCATION FOR STRATIFIED TELEPHONE SURVEY DESIGNS

Robert J. Casady, U. S. Bureau of Labor Statistics  
James M. Lepkowski, University of Michigan

Key Words: Random digit dialing, Telephone surveys

## 1. THE CURRENT STATUS OF TELEPHONE SURVEY DESIGNS

The two stage random digit dialing design for sampling telephone households, first proposed by Mitofsky (1970) and more fully developed by Waksberg (1978), has been widely employed in telephone surveys. The Mitofsky-Waksberg technique capitalizes on the fact that working residential numbers (hereafter referred to as WRNs) tend to be highly clustered within banks of consecutive telephone numbers. Currently, only about twenty percent of the possible telephone numbers within the known area code, three digit prefix combinations are WRNs. However, if a bank of 100 consecutive telephone numbers can be identified that has at least one known WRN then, on average, over 50 percent of the numbers in the bank will be WRNs. The Mitofsky-Waksberg technique, which identifies 100-banks containing WRNs in the first stage of sampling, greatly reduces the amount of screening necessary to identify telephone numbers assigned to households.

Alternatively, lists of published telephone numbers have been employed as sampling frames. These lists of published numbers are available for the entire country from commercial firms such as Donnelley Marketing Information Systems. A straightforward selection of telephone numbers from such lists provides a very high rate of WRNs (typically at least 85%) but unfortunately does not cover households with unpublished numbers. Comparisons of telephone households with and without published numbers (see, for example, Brunner and Brunner, 1971) indicates that substantial bias may result.

The purpose of this paper is to examine stratified designs based on the BellCore Research (BCR) frame as an alternative to list frames and Mitofsky-Waksberg design. As an example of frame stratification, the BCR frame could be partitioned into two strata: a "high density" stratum consisting of residential numbers in 100-banks with one or more listed numbers and a "low density" stratum consisting of all the remaining numbers in the BCR frame. Direct access to all listed numbers is not required for this stratification scheme. Counts of listed numbers, or any other indicator of the presence of listed telephone numbers in a 100-bank obtained from a reverse directory or a commercial list, would be sufficient. Preliminary work indicates that it is reasonable to expect that approximately 52% of the

numbers in the high density stratum will be WRNs, contrasted to only about 2% in the low density stratum. The obvious cost difference of sampling from the two strata can be exploited through differential sample allocation.

The next section examines the question of the appropriate allocation of sample between the strata when simple random sampling is utilized within each stratum. A key feature of the stratified telephone sample approach is that it permits alternative approaches to sample selection within the different strata. Several alternatives are presented and discussed in Section 3. The paper concludes with a general discussion contrasting the Mitofsky-Waksberg procedure and stratified designs.

## 2. THE ALLOCATION PROBLEM FOR STRATIFIED TELEPHONE DESIGNS

### 2.1 Background

For the purposes of this paper we will assume that the basic sampling frame is the collection of all telephone numbers generated by appending four digit suffixes to the BCR list of area-prefix codes. We will assume that each household in the target population is "linked" to one and only one telephone number in the basic sampling frame.

We will also assume that we have access (possibly only indirect) to a directory based, machine readable list of telephone numbers such as that available through Donnelley Marketing. It should be noted that because many households choose not to list their telephone numbers in a directory, any such directory based frame will not contain all of the WRNs. Also directory based lists are, by nature, out of date so they will omit some numbers that are currently WRNs while including others that are no longer WRNs.

From a survey design point of view these two frames tend to be radically different. The BCR frame includes all WRNs so it provides complete "coverage" of the households in the target population, but only about 20 percent of the telephone numbers included in the BCR frame are actually WRNs. Thus, the "hit rate" (and hence sampling efficiency) will be quite low for a simple RDD sample design utilizing the BCR frame. In contrast, a typical directory/list frame covers only about 70 percent of the target households, but the hit rate is 80 to 90 percent. In general the sampling efficiency for a simple RDD design using a directory/list frame is far better than can be attained for the BCR frame using the Mitofsky-Waksberg technique. Unfortunately, the low coverage rates associated with directory based frames preclude their use in many cases.

The basic idea of the proposed stratification approach is to utilize information from the directory based frame to partition the BCR frame into two or more strata with relatively high hit rates and a single stratum with a very low hit rate. The sample is allocated to the strata so as to minimize cost (variance) for a specified variance (cost). Hereafter the low hit rate stratum will be referred to as the residual stratum.

## 2.2 Basic Notation

Assume that the BCR frame of telephone numbers has been partitioned into  $H$  strata based on a 100-bank attribute which can be determined from the directory based frame. The choice of 100-banks is arbitrary, 10-banks or any other sized banks would work just as well. For the  $i^{\text{th}}$  stratum let

$P_i$  = proportion of the frame included in the stratum,

$h_i$  = proportion of the telephone numbers in the stratum that are WRNs (i.e., the hit rate),

$w_i$  = average proportion of WRNs in the non-empty 100-banks (i.e., the average hit rate for non-empty banks),

$z_i$  = proportion of the target population included in the stratum, and

$t_i$  = proportion of 100-banks in the stratum that contain no WRNs.

The average hit rate for the frame is  $\bar{h} = \sum_{i=1}^H h_i P_i$  and the proportion of empty 100-banks in the frame is

$\bar{t} = \sum_{i=1}^H t_i P_i$ . In general only the  $P_i$ 's will be known

with certainty. Data from a joint research project involving the Bureau of Labor Statistics and the University of Michigan were used to provide approximate values for the parameters  $h_i$  and  $w_i$  for the two strata in the example. Values for the remaining parameters were calculated using the

algebraic relationships  $t_i = 1 - (h_i/w_i)$  and  $z_i = \frac{h_i P_i}{\bar{h}}$ .

The approximations for all of the frame parameters for the two stratum design are given in Table 1. The values in the table imply that for the BCR frame  $\bar{h} \cong .211$  and  $\bar{t} \cong .605$ , which are in close agreement with those given in Waksberg (1978).

## 2.3 The Basic Estimation Problem, Sample Designs and Estimators

We assume the telephone numbers in the  $i^{\text{th}}$  stratum are labeled 1 through  $M_i$  and we let

$$d_{ij} = \begin{cases} 1 & \text{if the } j^{\text{th}} \text{ nb. in the } i^{\text{th}} \text{ stratum is a WRN,} \\ 0 & \text{otherwise.} \end{cases}$$

The variable of interest is the household characteristic  $Y$ , and  $y$  represents the value of  $Y$  for a particular household. The population parameter to be estimated is the population mean  $\mu = Y./N$ . where

$$N. = \sum_{i=1}^H \sum_{j=1}^{M_i} d_{ij} = \sum_{i=1}^H N_i \text{ and } Y. = \sum_{i=1}^H \sum_{j=1}^{M_i} d_{ij} y_{ij} . \text{ The term}$$

$N_i$  denotes the number of WRNs in the  $i^{\text{th}}$  stratum and  $N.$  denotes the number of WRNs in the population.

We denote simple random sampling without replacement (i.e., simple RDD) from the telephone numbers in the BCR frame as design  $D_0$  and stratified simple random sampling from the BCR frame (i.e. independent simple RDD samples are selected from each stratum) as design  $D_1$ . Under design  $D_0$  the standard ratio estimator for  $\mu$  is given by  $\bar{Y}_0 = \hat{Y}_0 / \hat{N}_0$  where  $\hat{Y}_0$  and  $\hat{N}_0$  are the usual inflation estimators for  $Y.$  and  $N.$  respectively. The estimator  $\bar{Y}_0$  is asymptotically unbiased for  $\mu$  and its variance is given by

$$\text{var}(\bar{Y}_0) \cong \frac{\sigma^2}{mh}$$

where  $m$  is the sample size and  $\sigma^2$  is the population variance of the  $y$ 's. For the design  $D_1$  the standard ratio estimator of  $\mu$  is given by  $\bar{Y}_1 = \hat{Y}_1 / \hat{N}_1$  where  $\hat{Y}_1$  and  $\hat{N}_1$  are the standard inflation estimators for  $Y.$  and  $N.$  under stratified sampling. The estimator  $\bar{Y}_1$  is also asymptotically unbiased for  $\mu$  and

$$\text{var}(\bar{Y}_1) \cong \sum_{i=1}^H \frac{z_i^2 \sigma_i^2 (1 + (1 - h_i) \lambda_i)}{m_i h_i} \quad (2.1)$$

where  $\lambda_i = (\mu_i - \mu)^2 / \sigma_i^2$  and  $m_i$ ,  $\mu_i$ , and  $\sigma_i^2$  are the stratum sample sizes, means, and variances, respectively.

## 2.4 The Cost Model

There are costs associated both with determining the value of the indicator variable  $d$  and the value of the characteristic of interest  $Y$ . The cost function for determining the indicator variable is denoted by  $C_1(\cdot)$ , with

$$C_1(d) = \begin{cases} c_1 & \text{if } d = 1 \\ c_0 & \text{if } d = 0 \end{cases}$$

This model allows for the possibility that the cost of determining that a telephone number is not a WRN may be different than determining that a telephone number is a WRN. In fact, the cost of determining the status of telephone numbers that are WRNs is usually less. The cost of determining the value of the characteristic  $Y$  includes only the additional cost of determining the value of  $y$  after the value of  $d$  has been determined. Accordingly, let  $C_2(\cdot, \cdot)$  represent this additional cost, then

$$C_2(d, y) = \begin{cases} 0 & \text{if } d = 0 \\ c_2 & \text{if } d = 1 \end{cases}.$$

The sum  $c_1 + c_2$  represents the cost of a "productive" sample selection and  $c_0$  represents the cost of an "unproductive" selection; following Waksberg (1978),  $\gamma = (c_1 + c_2) / c_0$  represents the ratio of the cost of a productive selection to an unproductive selection. The total cost for sample selection and the determination of the values of  $Y$  is a random variable for both designs. Letting  $C(D_0)$  and  $C(D_1)$  represent the total cost of conducting a survey under the two respective designs it is straightforward to show that

$$E[C(D_0)] = mc_0(1 + (\gamma - 1)\bar{h}) \quad (2.2)$$

and

$$E[C(D_1)] = c_0 \sum_{i=1}^H m_i (1 + (\gamma - 1)h_i). \quad (2.3)$$

## 2.5 Optimal Allocation for $\bar{Y}_1$

The stratum sample allocation that minimizes  $\text{var}(\bar{Y}_1)$  for a fixed expected total cost  $C^*$  (or that minimizes  $E[C(D_1)]$  for a fixed variance  $V^*$ ) is specified up to a proportionality constant by

$$m_i \propto \frac{z_i \sigma_i}{\sqrt{h_i}} \left( \frac{1 + (1 - h_i)\lambda_i}{1 + (\gamma - 1)h_i} \right)^{1/2} \quad (2.4)$$

where the proportionality constant is determined by substitution into the expected cost equation (or the variance equation, as appropriate). Relative to RDD sampling, the proportional reduction in variance (cost) under optimal allocation for fixed cost (variance), denoted by  $R(\bar{Y}_1, \bar{Y}_0)$ , is approximately

$$1 - \bar{h} \frac{\left[ \sum_{i=1}^H \frac{z_i \sigma_i}{\sqrt{h_i}} \left[ (1 + (1 - h_i)\lambda_i)(1 + (\gamma - 1)h_i) \right]^{1/2} \right]^2}{\sigma^2 (1 + (\gamma - 1)\bar{h})} \quad (2.5)$$

## 2.6 Practical Problems Associated With Optimal Allocation

The problem of specifying the values for the parameters in the allocation equations is generic to optimal allocation schemes. For our particular case there are three basic types of parameters: frame related ( $z_i$  and  $h_i$ ), cost related ( $\gamma$  and  $c_0$ ) and those specific to the variable of interest ( $\lambda_i$  and  $\sigma_i^2$ ). Currently, we have a fairly good working knowledge of the frame related parameters for the two stratum example and certain other specific stratification schemes. In Section 4, we will discuss several active research projects that should further expand our knowledge in this area.

It is clear that  $\gamma \geq 1$  but the actual value can vary widely. Waksberg (1978) considers values of  $\gamma$  between 2 and 20. Potentially the variable specific parameters pose the most serious problem. Usually our knowledge regarding the values of these parameters is limited and, in the case of multipurpose surveys, we must decide which variable(s) to use for the purposes of allocation. Fortunately, in many practical applications, two factors combine to somewhat lessen this problem. First, the allocation tends to be relatively "flat" in a neighborhood of the optimum allocation so that the reduction in variance is relatively robust with respect to allocation. Secondly, the variables of interest will not usually be highly related to variables we are using for stratification. Therefore, with caution, we assume that  $\mu_i = \mu$  and  $\sigma_i^2 = \sigma^2$  for  $i = 1, 2, \dots, H$ . Optimal allocation is achieved by

$$m_i \propto \frac{z_i}{\sqrt{h_i}} (1 + (\gamma - 1)h_i)^{-1/2} \quad (2.6)$$

and the proportional reduction in variance is

$$R(\bar{Y}_1, \bar{Y}_0) \cong 1 - \bar{h} \frac{\left[ \sum_{i=1}^H z_i \left( \frac{1 + (\gamma - 1)h_i}{h_i} \right)^{1/2} \right]^2}{(1 + (\gamma - 1)\bar{h})} \quad (2.7)$$

In the case of the two stratum example, the allocation specified by (2.6) implies that the allocation relative to the residual stratum (i.e.,  $m_1 / m_2$ ) is 2.54 when  $\gamma = 2$  and 1.42 when  $\gamma = 10$ . In the first case the projected proportional reduction in variance is  $R = .283$  and in the second  $R = .077$ . In fact, it follows from (2.7) that as the relative cost of determining the value of the variable of interest increases, the relative benefit of optimal allocation decreases.

The Mitofsky-Waksberg sample design, denoted by  $D_3$ , employs two stages of sample selection (i.e.,

non-empty 100-banks are selected in the first stage and WRNs are selected in the second stage). The Mitofsky-Waksberg estimator, denoted by  $\bar{Y}_3$ , is unbiased for  $\mu$ . Under "optimal" within 100-bank sample allocation, the reduction in variance relative to simple RDD for the estimator  $\bar{Y}_3$ , denoted by  $R$ , is approximately

$$1 - \frac{\left[ (1 + (\gamma - 1)\bar{h} - \bar{t})^{1/2} (1 - \rho)^{1/2} + (\rho\bar{t})^{1/2} \right]^2}{1 + (\gamma - 1)\bar{h}} \quad (2.8)$$

where  $\rho$  is intra-bank correlation. At the national level Groves (1977) reports that  $\rho \approx .05$  for economic or social statistics. Using this value of  $\rho$ , together with the values of  $\bar{h}$  and  $\bar{t}$  from the two stratum example, the projected proportional reduction in variance for the Mitofsky-Waksberg procedure is  $R = .281$  when  $\gamma = 2$  and  $R = .060$  when  $\gamma = 10$ .

The two methodologies appear to produce essentially identical variance reduction for both values of the cost ratio. However, too much should not be read into this simple comparison as the projected reduction for each of the procedures is based on simplifying assumptions that will not be strictly true for any application. The only inference intended is that the two procedures appear to highly competitive under a general set of circumstances typically encountered in application.

### 3. Alternative Sample Designs

#### 3.1 Truncated Designs

The designs presented in the previous section produce unbiased estimates of the population mean. Incorrect assumptions regarding the various frame, cost, and population parameters only effect the efficiency of the estimators, not their expectations. Unfortunately an extremely high price is paid for the assurance of unbiasedness because sampling from the residual stratum provides information on only a small proportion of the population and at a relatively high cost. If we are willing to settle for an estimate of the population mean exclusive of those households linked to telephone numbers in the residual stratum, we can "truncate" the original frame by eliminating the residual stratum and selecting a stratified RDD sample from the remaining telephone numbers. For the two stratum example the "truncated frame" would consist only of those telephone numbers in the first stratum. The hit rate for the sample from the truncated frame would be .521, in contrast to a hit rate of .211 for the entire frame. However, only about 94% of the target population would remain in scope.

In what follows we assume that the truncated frame is simply the original BCR frame less the residual stratum. Accordingly, for the truncated frame

$\bar{h}^* = (\bar{h} - P_K h_K) / (1 - P_K)$  is the hit rate, the proportion of empty 100-banks is  $\bar{t}^* = (\bar{t} - P_K t_K) / (1 - P_K)$  and  $\mu^* = (\mu - z_K \mu_K) / (1 - z_K)$  is the population mean. Let design  $D_4$  be stratified simple random sampling from the truncated frame, and  $\bar{Y}_4$  the standard ratio estimator of the population mean. The estimator  $\bar{Y}_4$  is asymptotically unbiased for  $\mu^*$ , and, in general, it is biased for  $\mu$ . The (asymptotic) bias is given by

$$B(\bar{Y}_4) = \mu^* - \mu = \frac{z_K (\mu - \mu_K)}{(1 - z_K)}. \quad (3.1)$$

In most practical circumstances the bias tends to zero monotonically as the proportion of the target population in the residual stratum becomes small, although, as indicated by (3.1), this is not necessarily the case. In any event, since the value of  $\mu - \mu_K$  is never known, an upper limit on the proportion of the population in the residual stratum is usually the key specification to be determined when considering the use of a truncated frame. For the two stratum example approximately 6% of the target population is excluded from the sampling frame and, in almost all cases, this would not be tolerable for Federal agencies.

The equations for cost, variance, allocation, and proportional reduction in variance (or cost) are essentially the same as those presented in Section 2. In fact, the only modifications required for equation (2.1) and equations (2.3) through (2.7) are to replace  $\mu$  with  $\mu^*$ ,  $z_i$  with  $z_i^* = z_i / (1 - z_K)$ , and  $\lambda_i$  with  $\lambda_i^* = (\mu_i - \mu^*)^2 / \sigma_i^2$  where  $i = 1, 2, \dots, H - 1$ . Obviously, all sums are only over the remaining  $H - 1$  strata. For the special case where only one stratum remains after truncation the proportional reduction in variance (cost) reduces to

$$R(\bar{Y}_4, \bar{Y}_0) = 1 - \frac{\bar{h}(1 + \bar{h}^*(\gamma - 1))}{\bar{h}^*(1 + \bar{h}(\gamma - 1))}. \quad (3.2)$$

Thus for the two stratum design, the proportional reduction in variance (cost) is approximately .49 when  $\gamma = 2$  and .21 when  $\gamma = 10$ . In both cases the reduction is substantially greater than achieved by the two methods in the previous section. However, nearly 6% of the population is not covered by the frame.

In an attempt to retain the relative efficiency of truncation while reducing the magnitude of the coverage problem, BLS and the University of Michigan are investigating several alternative stratification plans in an effort to reduce the proportion of the population in the residual stratum.

One promising approach calls for the partitioning of the residual stratum in the two stratum example to form a new residual stratum which consists of telephone numbers in 100-banks thought to be primarily assigned to commercial establishments or not yet activated for either residential or commercial use. Estimated frame parameters for the resulting three stratum design are given in Table 2.

These data were used to compute the projected proportional reduction in variance for both the three stratum design and the truncated three stratum design. These results, together with a summary of the results for the two stratum designs and the Mitofsky-Waksberg design, are presented in Table 3. Table 3 also includes the projected reduction in variance for a cost ratio of 20.

It appears that the proposed partitioning strategy was reasonably successful as the percent of the population out of scope was reduced from nearly 6% to approximately 2%. The projected proportional reduction in variance for the truncated three stratum design is approximately .41 when  $\gamma = 2$  and .16 when  $\gamma = 10$ . From an efficiency point of view, it occupies the middle ground between the highly efficient truncated two stratum design and unbiased designs. Of course the issue to be faced when considering such a design is the coverage problem. For any particular application the risk inherent in sampling from a frame that does not include all of the target population must be weighed against the potential gain in efficiency. As expected, the standard three stratum design is slightly more efficient than the two stratum design. However, the increase in efficiency is so small that it is doubtful that the added cost of partitioning the BCR frame into an additional stratum is justified except for the purpose of truncation.

### 3.2 Designs Using Both Optimal Allocation and the Mitofsky-Waksberg Procedure

The final design to be considered is based on the stratified BCR frame. Depending on the proportion of empty 100-banks in the stratum, we use simple RDD sampling in some strata and Mitofsky-Waksberg sampling in others. The motivation for this type of design is based on the following two considerations:

- (a) Mitofsky-Waksberg sampling tends to be "administratively complex", and if the gain in efficiency is small, simple RDD is preferred.
- (b) If the proportion of empty banks in a stratum is "small" then Mitofsky-Waksberg sampling offers little, if any, increase in efficiency.

Thus, we propose to utilize simple RDD sampling in strata with a "small" proportion of empty hundred banks and Mitofsky-Waksberg sampling in the remaining strata. A complete discussion of this topic may be found in Casady and Lepkowski (1991).

## 4. CONCLUDING REMARKS

The strengths of the Mitofsky-Waksberg technique for generating telephone samples are clear: high hit rates in the second stage of selection, an efficient method for screening out empty banks of telephone numbers, and a conceptually ingenious approach to sample generation. It is a remarkable testimony to the strength of the technique that after many years it is still considered to be the standard method of random digit dialing with few serious competitors. The weaknesses of the technique (first stage screening and replacement of non-residential numbers during the data collection) do not, on the surface, seem to be important relative to the general strength of the technique. However, these features can cause substantial difficulty, especially in short time period telephone survey operations.

Several alternatives are considered. In the two stratum methods, telephone numbers are divided into two groups, the high density stratum consisting of all telephone numbers in 100-banks containing listed numbers and the low density stratum consisting of all remaining telephone numbers. In the three stratum methods, the "all other numbers" or original low density stratum is further subdivided, on the basis of auxiliary data, into two strata. One of the new strata is expected to contain nearly all residential numbers in the original low density stratum and the other is now the (new) low density stratum. For both the two and three stratum design, two general alternatives are considered: (1) selecting simple random samples from all strata except the low density stratum frame where the Mitofsky-Waksberg method is used and (2) selecting simple random samples from all strata except the low density stratum which is not sampled at all.

The thesis of this paper is that alternative telephone sampling methods using counts of listed numbers are available that are at least as efficient as the Mitofsky-Waksberg technique. Auxiliary information from listed number frames are used, rather than a random digit dialing screening technique, to screen empty banks of numbers. Further, the use of these designs can eliminate the need for the troublesome replacement of non-residential numbers at the second stage, since the only telephone numbers that must be dialed in the high density stratum are those that are generated at the beginning of the study.

Even for low cost ratios, the two and three stratum designs are as efficient as the Mitofsky-Waksberg approach. When numbers can be dropped from the low density stratum, these alternatives are much more efficient, at the price of unknown bias due to excluding part of the target population. When cost ratios are high, the two and three stratum approaches are clearly superior.

A critical issue is the magnitude of the bias introduced by dropping the low density stratum. Even though the proportion of the population in the stratum is small, the magnitude of the bias may be relatively large for some characteristics and for some subgroups of the population. Further empirical investigations are necessary.

The cost of the auxiliary list frame is not addressed in this investigation because the frame information used in stratification was derived from a specialized research file. Further investigation is needed into this cost as it must be considered in any practical application.

In order to improve the hit rates in the high density stratum, smaller banks of numbers can be used. In another investigation we have found that 10-banks will have hit rates in the neighborhood of .57 compared to the .52 reported here for 100-banks. Working with 10-banks substantially increases the size of files and processing operations used to generate samples. And, the cost of a 10-bank frame is likely to be much higher than the 100-bank frame.

Clearly the results presented here are insufficient to draw final conclusions about the overall value of

these alternative designs. Further cost data and empirical evidence on the size of the bias caused by eliminating the numbers from the low density stratum is needed before a final conclusion can be reached.

## 5. References

- Brunner, J. A., and Brunner, G. A., "Are Voluntarily Unlisted Telephone Subscribers Really Different?" *Journal of Marketing Research*, Vol. 8, February 1971, pp.121-124.
- Casady, R. J. and Lepkowski, J. M., "Stratified Telephone Designs," unpublished report of the U. S. Bureau of Labor Statistics, Washington D.C.
- Groves, R. M., "An Empirical Comparison of Two Telephone Sample Designs," unpublished report of the Survey Research Center of the University of Michigan, Ann Arbor, MI.
- Mitofsky, W., "Sampling of Telephone Households," unpublished CBS News memorandum, 1970.
- Waksberg, J., "Sampling Methods for Random Digit Dialing," *Journal of the American Statistical Association*, Vol. 73, No. 361, March 1978, pp.40-46.

**Table 1.** Approximate values of the frame parameters for a two stratum design based on the BCR frame and the Donnelley list frame.

|   | Proportion of Frame ( $P_i$ ) | Proportion of Population ( $z_i$ ) | Hit Rate ( $h_i$ ) | Proportion of Empty 100-Banks ( $t_i$ ) | Hit Rate Within Non-empty Banks ( $w_i$ ) |
|---|-------------------------------|------------------------------------|--------------------|-----------------------------------------|-------------------------------------------|
| 1 | .3804                         | .9402                              | .5210              | .0300                                   | .5371                                     |
| 2 | .6196                         | .0598                              | .0204              | .9584                                   | .4900                                     |

**Table 2.** Estimated frame parameters for a proposed three stratum design based on the BCR frame and the Donnelley list frame.

| Stratum | Proportion of Frame ( $P_i$ ) | Proportion of Population ( $z_i$ ) | Hit Rate ( $h_i$ ) | Proportion of Empty 100-Banks ( $t_i$ ) | Hit Rate Within Non-empty Banks ( $w_i$ ) |
|---------|-------------------------------|------------------------------------|--------------------|-----------------------------------------|-------------------------------------------|
| 1       | .3804                         | .9402                              | .5210              | .0300                                   | .5371                                     |
| 2       | .2000                         | .0399                              | .0420              | .9143                                   | .4900                                     |
| 3       | .4196                         | .0199                              | .0100              | .9796                                   | .4900                                     |

**Table 3.** Projected proportional reduction in variance (or cost) relative to simple RDD sampling for five alternative telephone sample designs.

| Sample Design             | Proportional Reduction in Variance or Cost |               |               | Proportion of Frame Not in Scope |
|---------------------------|--------------------------------------------|---------------|---------------|----------------------------------|
|                           | $\gamma = 2$                               | $\gamma = 10$ | $\gamma = 20$ |                                  |
| Two Stratum               | .2829                                      | .0766         | .0320         | .0000                            |
| Two Stratum (Truncated)   | .4917                                      | .2055         | .1189         | .0598                            |
| Mitofsky-Waksberg         | .2811                                      | .0597         | .0135         | .0000                            |
| Three Stratum             | .3001                                      | .0866         | .0389         | .0000                            |
| Three Stratum (Truncated) | .4095                                      | .1574         | .0879         | .0199                            |