A List-Assisted Telephone Sampling: Conjugate Pair Selection.

Hee-Choon Shin

NORC, 55 East Monroe St., Suite 4800, Chicago, IL 60603

KEY WORDS: Mitofsky-Waksberg Procedure, RDD, Efficiency, Bias, Conjugate Pairs

Introduction

In conducting telephone surveys, the random digit dialing (RDD) method is the best approach to obtain unbiased estimates from the population with telephone services. The Mitofsky-Waksberg (M-W) technique and modified versions of M-W techniques has been applied to the majority of large-scale RDD surveys. The most difficult part of a RDD survey is to locate households through working residential numbers (WRNs).

This paper proposes an efficient method of increasing the WRN rate while minimizing the potential bias of estimates.

Background

Typically, at least in contemporary U.S.A., a relatively pure form of RDD sample consists of a set of randomly generated 10-digit phone numbers within the active exchanges. The sampling frame includes all the possible residential numbers. i.e., telephone households. The WRN rate of a sample from this population is approximately 19-20 %, depending on the number of working 100-banks. An earlier attempt to increase the WRN rate was the M-W approach (Waksberg, 1978). Further, with the availability of commercial lists of residential phone numbers, various forms of list-assisted RDD methods were developed. The most common approach is to restrict the sample selection only to the working 100-banks with 1 or more of listed households, or 1+ RDD method. A 40% of WRN rate is expected from the 1+ RDD method. The next logical but undesirable approach to increase the WRN rate is to select sample telephone numbers from the listed households with telephone services, i.e., list sample approach. The problem of the list sample approach is in the fact that a substantial portion of residential households are not listed in the sampling frame, and there are serious concerns about the potential for bias.

Alternative approaches have been suggested to compensate for the lower WRN rates of the RDD methods and the biasedness of the list sample approach. Stratified RDD methods and dual frame approaches were proposed and used (Casady and Lepkowski, 1993). The weakness of these approaches is the presence of extremely small probability of selection from RDD stratum or RDD portion of the sample. Higher weighting effects negate the benefits of stratification, and result in a smaller effective sample size (Elliott and Little, 2000).

In the following, we present a hybrid approach between 1+ RDD and list sample approaches, and propose an alternative to the stratification of the frame. Our aim is to increase the WRN rate of the sample while maintaining unbiasedness and efficiency of the estimates.

Conjugate Pair Selection

The essence of our proposed approach is in selecting n pairs of phone numbers within each bank. We choose n pairs of numbers in following way. Consider a pair of telephone numbers:

1) Choose a phone number from the listed WRNs and denote its last two digits as x.
2) The other number of the pair which will be included in the sample is x' where $x + x' = 99$.

We call this ordered pair of (x, x') a conjugate pair. Let $x = a + bi$ where i denotes $\sqrt{-1}$ in C, the complex field. Then, the conjugate of $x = a + bi$ is $x' = a - bi$, or *vice versa*.

Given $x + x' = 99$, the following relationships follow:

$$a = \frac{1}{2}\left|x + x'\right|$$

$$b = \frac{1}{2}\left|x - x'\right| i$$

As stated, the term x is the last two digits of a WRN. However, the term x' could be 1) a WRN, 2) an unlisted WRN, or 3) other type of numbers including disconnected and business numbers.

Note that x is randomly selected from the listed WRNs within a 100-bank but x' is a function of x. Therefore, the resulting sample of telephone numbers is a self-weighting equal probability sample except the duplicity in case of the pair with two listed WRNs.

Naturally the minimum of WRN rate of the sample generated in this fashion is 50%.[1]

Consider Table 1 for a classification of possible pairs.

Our approach will cover WRNs from $(\eta_{11}, \eta_{12}, \eta_{21}, \eta_{13}, \eta_{31})$ of pairs. Since we are interested in finding WRNs in each pair, we can ignore phone numbers in η_{33}. The WRNs in $(\eta_{22}, \eta_{23}, \eta_{32})$ are the possible sources of bias. The viability of our approach depends on the following two factors: 1) the ratio of the WRNs in these 3 cells to the total number of WRNs; 2) size of bias.

Initial List Sample: Two Approaches

In the following, we describe two specific approaches in obtaining an initial sample of listed households for conjugate pairs. Using this initial sample, the second sets of numbers will be generated using the functional relationship shown above.

Two-stage approach.

First, a set of m banks is selected with probability proportional to size from the frame with M banks. The size, MOS_i of the ith bank is the number of listed WRNs in the bank. The selection probability of the ith bank is

$$P_i = m\frac{MOS_i}{\sum_{i=1}^{M} MOS_i}.$$

Second, select n_i WRNs from each of selected ith bank. Given the ith bank, the conditional selection probability of the jth phone number is

$$P_{j\backslash i} = \frac{n_i}{MOS_i}.$$

Then the unconditional selection probability of the jth phone number within the ith bank is

$$P_{ij} = P_i P_{j\backslash i} = m\frac{MOS_i}{\sum_{i=1}^{M} MOS_i}\frac{n_i}{MOS_i} = \frac{mn_i}{\sum_{i=1}^{M} MOS_i}$$

To generate a self-weighting sample of WRNs, n_i should be a constant, n. Unless we restrict the sampling to very small area, usually n is 1.

Single-stage approach.

The number of mn initial sample of WRNs can be directly selected from the frame (the list of known WRNs). The frame could be sorted by exchange and bank indicators, and the sample could be systematically selected from the list frame. The selection probability is simply

$$P_{ij} = \frac{mn}{N_L},$$

where N_L is the frame size.

Inclusion Probability of a Phone Number

Each phone number would be included in the sample in 2 ways: 1) As a selected number in the initial sample of listed WRNs, and 2) as a conjugate number of a pair. Let $P(x)$ be the selection probability of initial sample of listed WRNs within each bank. Then the inclusion probability of each pair is:

$$\begin{aligned} P\left(x \in Listed \cap x' \in Unlisted\right) &= P(x)\cdot P(x' \backslash x) \\ &= P(x) \end{aligned}$$

[1] We assume that the list frame is complete without any errors or omissions. In reality, the proportion of WRNs in the list frame is less than 100%.

$$P\left(x \in Listed \cap x^{'} \in Listed\right)$$
$$= P(x) \cdot P(x^{'} \setminus x) + P(x^{'}) \cdot P(x \setminus x^{'})$$
$$= 2P(x)$$

The inclusion probability of phone numbers in list-list pair is twice that in list-unlist or unlist-list pair.

Simulation

Figure 1 shows the distribution of 100-banks with 1 or more of listed households by the number of listed households.[2] Currently there are about 2.7 million such banks. The average number of listed households in all the 100-banks with 1 or more of listed households is 26.19. That is, of all the possible telephone numbers in 1+ banks, 26.19% are listed households. The graph seems to indicate that a set of new working banks were recently designated. By subtracting the number of listed households from the total telephone households, it is estimated that 13.33% are unlisted households.[3] The existence of unlisted WRNs is the reason for the use of RDD sampling. A RDD approach attempts to give a non-zero chance of selection to unlisted households.

To measure the effectiveness and the efficiency of our approach, we conducted a simulation study in the following way:

1) Generate a population of 100-banks with the distribution of listed households shown in Figure 1.
2) Generate unlisted households from Poisson distribution with mean of 13.33 within each bank.[4]
3) Select the initial sample of listed households
4) Find conjugate pairs

We examined the following 3 statistics to measure the desirability of our proposed approach.

1) Proportion of WRNs in the sample,
2) Proportion of unlisted WRNs in the sample,
3) WEFF – weighting effect, or Design effect (DEFF) due to the unequal inclusion probabilities, which is approximated by $1 + CV^2$. The CV is the coefficient of variation of the sampling weights.

Discussion

Table 2 shows our simulation strategies and corresponding results. Let us look at the last row as a specific example: We assumed that the number of telephone numbers in the universe was 1 million with listed households, unlisted households, and other non-residential numbers. We selected an initial sample of 1,000 listed numbers. We repeat the whole procedures 100 times. The average proportion of WRNs in the simulated samples is 73.93%. In the samples, the average proportion of unlisted households was 6.68%, which was about half of unlisted households. The weighting effect was ignorable as indicated by WEFF=1.0000.

Overall, we could achieve about 73-74% of WRN rate while covering about one half of unlisted numbers by adopting our approach. An even higher hit rate would be obtained if we prescreen the resulting samples for business and disconnected numbers.

In summary, the conjugate pair approach increases the WRN rate to about 74%, covering about 50% of the unlisted households, while achieving a $DEFF \cong 1.00$. Given the inefficiencies of a "traditional" RDD, the high *DEFF*s of a stratified RDD approach, and the bias problem of a list sample, the conjugate pair approach seems promising and worthy of applying to a real-world telephone sample.

[2] Figure 1 is based on the statistics generated by a well-known commercial telephone sample vendor.

[3] According to the March Current Population Survey (2004), the estimated number of telephone households is 1.0556×10^8, or 94.15% of U.S. households.

[4] The Poisson distribution is a discrete function. For this simulation, we simulated the unlisted households from the Poisson distribution with mean of 13 in 33% of banks, and with mean of 14 in 67% of banks in each replication.

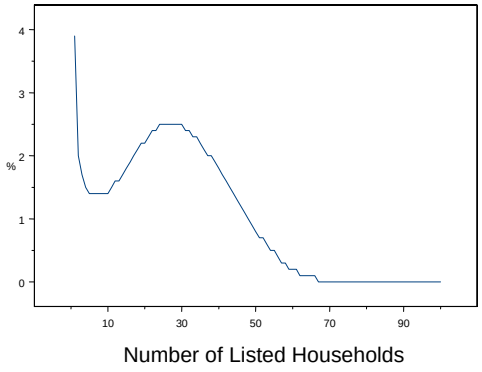
Table 1. A Classification of Possible Pairs

	LW	UW	UN
Listed WRN (LW)	η_{11}	η_{12}	η_{13}
Unlisted WRN (UW)	η_{21}	η_{22}	η_{23}
Unlisted Non-WRN (UN)	η_{31}	η_{32}	η_{33}

Table 2. Simulation Results

Pop. Size	List Sample Size	Reps.	Sample Characteristics		
			WRN Rate	Un-listed	WEFF
$1x10^4$	500	10	73.36%	6.69%	1.0000
	500	100	73.20%	6.94%	1.0000
	1,000	10	71.99%	6.85%	1.0000
	1,000	100	72.14%	7.17%	1.0000
$1x10^5$	500	10	73.60%	6.55%	1.0000
	500	100	73.98%	6.80%	1.0000
	1,000	10	73.80%	6.72%	1.0000
	1,000	100	73.81%	6.73%	1.0000
$1x10^6$	500	10	73.54%	6.26%	1.0000
	500	100	74.05%	6.78%	1.0000
	1,000	10	74.04%	6.74%	1.0000
	1,000	100	73.93%	6.68%	1.0000

Figure 1. Distribution of 100-Banks with 1 or more of Listed Households by the Number of Listed Households (As of April, 2005).



References

Brick, J.M, J. Waksberg, D. Kulp, and A. Starer. (1995), "Bias in list-assisted telephone samples," *Public Opinion Quarterly* **59**: 218-235.

Casady, R. and J. Lepkowski. (1993), "Stratified Telephone Survey Designs," *Survey Methodology* **19**: 103-113.

Elliott, M.R. and R.J.A. Little. (2000), "Model-Based Alternatives to Trimming Survey Weights," *Journal of Official Statistics* **16**: 191-209.

Waksberg, J. (1978), "Sampling Methods for Random Digit Dialing," *Journal of the American Statistical Association*. **73**: 40-46.

Acknowledgments

I am grateful to Javier Porras for his valuable comments.