ABSTRACT: A Refined LARDD Design in the PR Telephone Points of Purchase Survey

The Puerto Rico Consumer Expenditure Survey included a telephone points of purchase survey as part of the consumer price index revision. The US Bureau of Labor Statistics collaborated with the PR Department of Labor in technical assistance at different phases. An electronic file of all listed residential telephone numbers, as of April 20, 2001, and the prevalence of unlisted numbers in 100-numbers non zero clusters, or "unlisted clusters", were used to refine random digit dialing to generate the telephone sample.
This was a first-time approach to increase frame efficiency and reduce implementation complexities in the PR telephone surveys. A pps cluster sample was selected from unlisted clusters with, at least both, 15 % hit rate and 10 unlisted numbers; systematic sampling was used in the listed numbers frame. Six pseudo replicates (waves) per Commodity Group (POPS) were selected during a 6-month period, stratified by geographic region, resulting in 44,603 initial calls and 8,524 final interviews.

Introduction

Aproximately 75% of Puerto Rican households have telephones (estimates based on U.S. Bureau of the Census household count, 2000); close to 30% of those are private or non-listed residential numbers (NLR). During the last decade telephone surveys in Puerto Rico (PR) have begun to receive serious consideration as efficient and valid alternatives for data collection. Exploring ways to improve process administration and increase frame efficiency is thus warranted. Evidence supports that even though other factors may grow in importance, list assisted random digit dialing methods (LARDD) are increasingly efficient in light of recent telephone system changes. (Tucker, Lepkowski and Piekarsi, 2002).

In 2001, LARDD was implemented in PR in The Telephone Points of Purchase Survey (TPOPSPR). This survey follows the U.S. Bureau of Labor Statistics TPOPS survey designed to generate an establishment sampling frame for the U.S. Consumer Price Index, based on consumer shopping patterns reported in telephone interviews. The PR Department of Labor and Human Resources sponsored the TPOPSPR survey,

with the technical collaboration of the U.S. BLS. The BLS - CATI questionnaire was translated to Spanish and adapted to the PR idiosyncracies.

This paper describes the refined LARDD approach utilized to generate the telephone sample. Expected and actual results for the RDD per bank approach are compared. Resulting gains in frame efficiency strengthen the need to improve access to list information. In the PR case, exploring the extent of the gains is important since obtaining additional information requires better communication channels between government agencies and a new model for conceptualizing the production of social and economic statistics in the public sector.

Survey Methodology

Constructing the Telephone Sampling Frame

An efficient telephone sampling frame is a primary goal in LARDD methodology. Not less desirable is simplifying operations: controlling under and over- identification of eligible numbers per bank; and, maintaining an effective call scheduling and follow up. Further, because of lack of sufficient information, RDD designs are associated with a two-stage approach in which a first number selection is used to classify the cluster as a non zero listed number bank or not; numbers are generated in the second stage from the non zero clusters. Thus, improving available information in LARDD is crucial for cost containment and time constraints.

At the time of the study, no significant experience had been accumulated in PR in preparing an efficient telephone sampling frame in electronic form. RDD approaches with limited auxiliary information had been used in some Marketing Research efforts. In this context, telephone numbers are usually generated randomly in a quota sampling

approach in which results are reported based on response households, with no follow up of not at homes or of "answering machine households". Little, if any, time period blocking for follow up calls is used. Difficulties also arise since cellular phones have increased dramatically, as well as households with unlisted phones.

Given the importance in TPOPSPR of cost considerations, time to completion and implementation ease, great efforts were directed at obtaining from the PR Telephone Company electronic files with significant auxiliary information to impact frame efficiency . Thus, the PR Telephone Co. provided an electronic listed residential frame (LRN) suitable for direct sampling of listed numbers, as well as per bank information for residential banks. Banks were identified by the telephone company as residential in the population of interest. Cellular prefixes were eliminated.

Originally 23,800 banks were identified as residential, of which additional information signaled the 18,596 with at least one non listed residential number So, identification of potential banks suppliers of residential working numbers was achieved via information supplied by the telephone company.

Figure 1 shows the distribution of non listed numbers per bank; a heavy left tail distribution with mean number of 16.6 non listed numbers per bank and a median of 17; Close to 17% of clusters having less than 10 working non listed numbers and only close to 13% more than 24.

Figure1
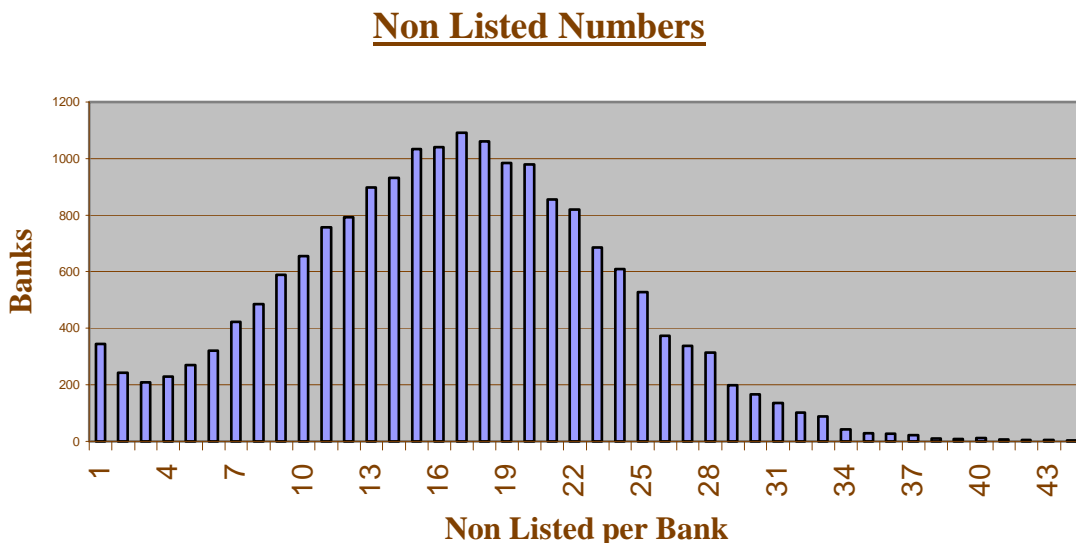
Non Listed Numbers

With such low density per bank, a Hit Rate measure was determined for each bank$_j$ based on non listed numbers density after eliminating listed numbers for each bank. Such an elimination increased search efficiency. For each bank the hit rate was:

$$\text{HIT RATE}_j \quad = \text{HR}_j = \quad \frac{\text{non listed working residential }_j}{100 - \text{Listed working residential}_j}$$

Figure 2 presents Hit Rates for the original 23,800 banks classified by the PR Telephone Co. as residential banks in the population of interest; the mean and median Hit Rate close to 0.28 for the non zero banks.

Figure 2

HIT   RATE   Based on Non Listed
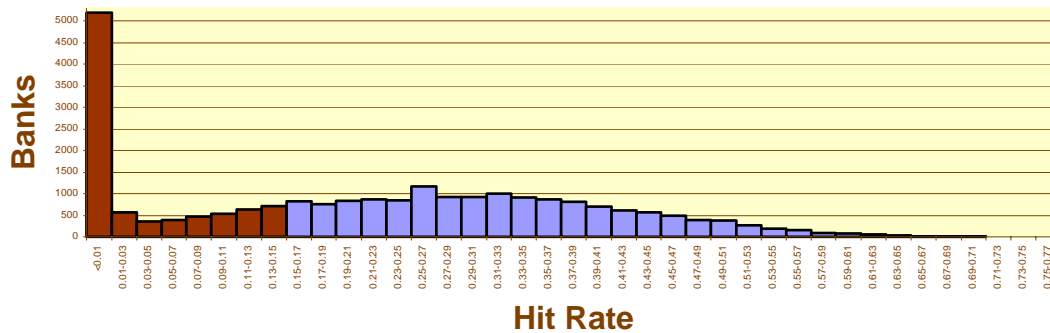
Frame Construction

A multiple frame approach was designed: the listed frame for direct sampling, and a list assisted frame to generate non listed numbers and new listed in a random digit dialing approach. The final telephone frame was developed as follows:

1. The electronic LRN file was stratified by geographical area; systematic random sampling generated a first sample of listed numbers

2. The Theoretical frame was stratified by 100-bank HIT RATE for selecting the NLR sample;   low HIT RATE banks were identified and truncated (those with less than 10 non listed numbers and hit rate less than 0.15);

3. Banks were ordered geographically and selected with pps based on NLR bank density;

4. Listed Numbers were eliminated from selected banks.

RDD process individualized by bank

The electronic files obtained permitted to preassign the number of digits to dial per bank for RDD implementation. RDD was used in the second stage of the NLR sampling

process after selecting banks with probability propotional to the number of non listed telephones.-

For each bank $_j$ the number of digits to be generated was preassigned so that the expected value was 10 working non listed numbers per bank (rounding increased the actual average to 11). RDD generated g_j digits per bank, where:

$$g_j \quad = \quad 10/ (0.85 \ HR_j)$$

Once this number was determined each telephone tried would be followed up until it could be classified as non response, interview, non eligible. Some remained unclassified. An adjustment of 0.85 times the hit rate was used to provide for information deficiencies.

Results

The sample was selected in 6 equivalent samples or Waves across 6 months. The results shown in Figure 3 are for Wave 5. Figure 3 shows the number of preassigned random digits to be generated per bank. In Wave 5 a minimum of 18 and a maximum of 61 random digits were generated in 136 banks selected with pps; an average of 37.8 per bank. Figure 4 in turn shows the resulting working non listed numbers per bank, an average of aproximately 11 working non listed numbers, with about 70% of banks having between 9 and 13 eligible numbers. Planned expected eligible numbers for Wave 5 was 10.9.
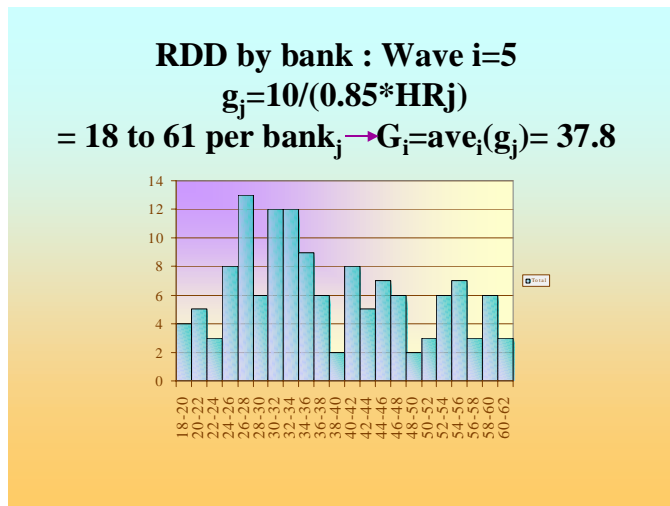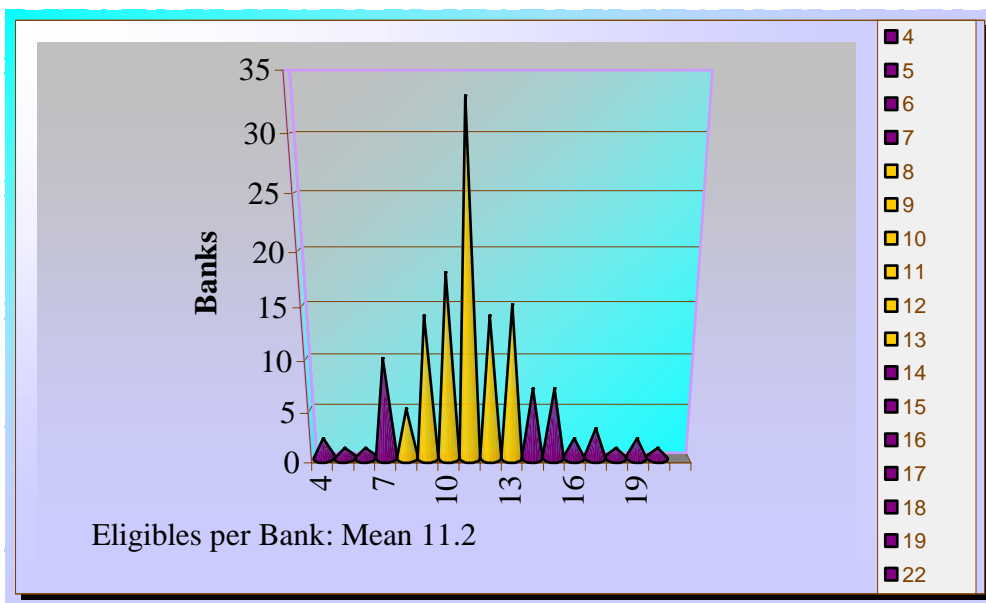
Figure 3

Figure 4 Number of Working Non listed Numbers per Bank

Conclusions

Frame efficiency can be notably improved with new auxiliary information complementing standard list assisted random digit dialing procedures. Cost containment and time restrictions still point to truncating the frame because of very low density banks.

A per bank approach provides a less cumbersome administration process; and, greater control of under and over representation of banks.

References

Tucker, C., Lepkowski, J.,Piekarsi,L. 2002. "The Current Efficiency of List Assisted Telephone Sampling Designs. " *Public Opinion Quarterly* 66: 321-338.