COSTS AND BENEFITS OF A PERMIT SAMPLE LATE IN THE DECADE

David Judkins, Diane Cadell, and Kelly Sczerba, Westat
David Judkins, Westat, 1650 Research Boulevard, Rockville, Maryland 20850

Key Words: Sample Design

1. Introduction

For household surveys, it is generally accepted that area sampling techniques coupled with face-to-face interviewing provide the lowest measurement biases of any design. This is due partly to the high coverage associated with preparing new listings of residential structures and partly due to the higher response rates that are associated with personal interviewing. However, such surveys are costly compared to techniques with higher biases such as random digit dialing and list sampling. The cost of area sampling can be sharply reduced with appropriate clustering and selection of clusters with probability proportionate to size. If good size measures exist, then these designs can be very efficient. However, the only technique that has been demonstrated to be practical in a national sample of primary sampling units is to use the Decennial Census. Of course, the Decennial Census is not available for processing until early in the year 01 following the census year. So the measures of size are much better for a survey conducted in 02 than 01, with a continual deterioration between these extremes. Samples in years 07 through 01 can have large design effects due to the deterioration in the measures of size.

To counter this deterioration, Joseph Waksberg developed a technique called permit sampling while at the Bureau of the Census. It was fully incorporated into the design of the Current Population Survey (CPS) by March of 1963 in conjunction with list sampling from the previous census (Technical paper 40, page 3, and Technical paper 7, page 69) and has been a feature of most ongoing demographic surveys at the Bureau since then. From Technical paper 7, we have the primary motivation for this type of sampling:

- "The main purpose in introducing address sampling from the 1960 Census, supplemented by sampled of permits for new construction, is to reduce the variance in segment size. This variance is sometimes quite large when areas sample segments are used and new construction is not adequately identified in a separate stratum." (Technical paper 7, page 69.)

Enlarging on this, we have from Waksberg (1998):

- "List sampling for household surveys was introduced in the CPS and other household surveys in the early 1960s, and I was responsible for that. As in most other developments, it was designed to overcome specific problems in area sampling. One of the problems was the quality of the maps. The second problem was the speed with which measures of size of segments based on the Census deteriorate. The 1950s and 1960s were periods of vast suburbanization of America, and the geography was changing rapidly. I remember that at one time, one of the segments in the CPS has grown so rapidly that if unbiased weighting has been used, it would have accounted for one-third of the African-American population of the United States."

Permit sampling is more attractive when used in conjunction with list sampling than in conjunction with area sampling because that way units built since the last census only have one change of selection. When permit sampling is used in conjunction with area sampling, the total costs are higher since the post-census housing is selected twice and must be screened out of the area sample. Nonetheless, permit sampling is also used in areas where list sampling could not be done due to the incompleteness of the census address books. Obviously, the quality of Census address books has improved dramatically since 1960, but outside contractors do not have access to those books for sampling, so the primary focus of this paper is on using permit sampling in conjunction with area sampling.

Unfortunately, little hard evidence of the benefits of permit sampling has ever been documented. Regarding benefits, Technical paper 40 (page 3) mentions that, "These changes resulted in a further gain in reliability of about 5 percent for most statistics." However, the proof of this statement is not offered and there is some ambiguity about whether the improvement was also partially due to the increase in the number of PSUs that year from 333 to 356. Some research was probably done prior to the implementation of the procedure in 1963, but research in BLS archives has failed to turn up any documentation. One of the current authors participated in some research at the Census Bureau in the early 1980s, but this effort was not very useful and the partial results were never submitted for publication. The following quote from the Gordon Report of 1962 (a report by the President's Committee to Appraise Employment and Unemployment Statistics) explains the difficulties in conducting the type of research that are still true today.

An overall review of the within-PSU sampling procedures, both existing and contemplated, cannot fail to impress one with the care that has been devoted to this phase of the survey design. Nevertheless, questions have been raised concerning the desirability of pushing this work still further. Unfortunately, objective analyses do not presently exist with which to demonstrate that the additional gains in precision that might be expected from such refinements would or would not be worth the expenditure of additional resources, particularly when there would be competing demands for the resources. The data for making the required studies exist in the monthly survey results, but, as noted previously, severe difficulties have been encountered in breaking into a complex data-processing operation that is geared to a tight time schedule and a more or less fixed budget. It is imperative that funds and computer time be made available for conducting studies of the type. (Gordon Report, Appendix D, pages 295-6, written by Philip J. McCarthy.)

On the other hand, the disadvantages of the procedure are fairly well known. (See for example Montaquila et al, 1999.) The technique is very complicated to implement and is often unpopular with survey administrators. A good overview of how to apply the method is given in Bell, et al (1999). Also, it introduces a variety of biases that are avoided with a pure oversampling approach. Thus, it has only been used at Westat for surveys conducted late in the 10-year redesign cycle. Up until now, the decision on whether to use a permit sample has been based on rough intuition and institutional traditions.

For this paper, we have taken advantage of a recent large listing effort for the National Survey of Parents and Youth (NSPY) to try to help survey researchers make more rigorous decisions about the desirability of permit sampling in new surveys. To this end, information was collected about the listing process in a fresh listing of about 3500 area segments in the Fall of 1999. Also information was collected about the costs of the sampling process for a permit sample in 135 building permit offices (BPOs) across 90 PSUs. The critical information from the area listing concerns the frequency and amplitude of new construction spikes. By its nature, much new construction is strongly clustered. It is these spikes that cause the efficiency of an area sample to deteriorate over the 10-year design cycle. Really large spikes are rare but devastating to survey precision when they occur, as indicated in the 1998 Waksberg quote above. Because of their rarity, their frequency can be well measured only in a survey with a very large number of segments. A sample of 3500 segments qualifies as very large. From the sampling of BPOs, the critical information is the cost. Here, it is important to have a staff that is working exclusively on the BPO listing in order to be able to get a good cost estimate from payroll records. This was the case on the NSPY BPO listing project.

In the rest of this paper, we review the theory on the effects of deterioration in the measure of size on survey efficiency, give the details of the deterioration measured in the NSPY area listing project, provide some cost information from the NSPY BPO listing project, and finally compare these costs.

2. Theory of Deterioration in Segment Measures of Size

Let M_{ti} be the number of dwelling units in segment I in year t after the Decennial Census. Let $\bar{n}$ be the desired number of sample dwelling units per segment. Let g be a small constant reflecting the maximum tolerance for growth in the sample size above the planned size for a segment. Then the sample size selected within a sample segment at year t (if the segment is selected) is

$n_i = min\left(\bar{n}g, \frac{M_{ti}}{M_{oi}}\bar{n}\right)$, and the probability of selection for the segment will be

$$p_i = \frac{M_{0i}}{c},$$

where c is some constant.

The Horwitz-Thompsen weight for sample dwelling units within the segment will be

$$w_{ij} = \frac{M_{ti}}{n_i}\frac{c}{M_{0i}}.$$

If the growth never exceeds a factor of g, then this weight will simplify to a constant $w_{ij} \equiv \frac{c}{\bar{n}}$. If growth is sometimes larger than a factor of g, then there will be a contribution to the design effect that can be approximated with the standard formula from Kish (formula 11.7.6 on p. 430 of Kish, 1965). This formula can be re-expressed so that the design effect is equal to 1 plus the relative variation in the weights. The relative variance in the weights is

$$V_w^2 = \frac{\sum_i n_i (\overline{w}_i - \overline{w})^2}{n\overline{w}^2},$$

where

$$\overline{w} = \frac{\sum_i n_i \overline{w}_i}{n}.$$

However, even if the weights are all equal, Kish's formula does not give the full design effect. There is also a consequence of unequal cluster sizes, but a good approximation of the effects of unequal cluster sizes that is easy to calculate has never been developed. Some approximations involve the relative variance in cluster sizes. This can be estimates as

$$V_n^2 = \frac{\sum_i (n_i - \overline{n})^2}{k\overline{n}^2},$$

where k is the number of sample segments.

One formula for the overall design effect is

$$DE = [1 + \rho(\overline{n} - 1)](1 + V_w^2 + V_n^2).$$

We explored potential design NSPY design effects using this approximation.

Other possibilities include:

$DE_2 = 1 + \rho(\overline{n} - 1) + V_w^2 + V_n^2$ from Waksberg and Northrup (1985),

$DE_3 = \{1 + \rho[\overline{n}(1 + V_n^2) - 1]\}(1 + V_w^2)$ from Verma (1993), and

$DE_4 = \{1 + \rho(\overline{n} - 1)\}(1 + V_w^2)$ from Kish (1987).

One of the advantage of the formula we used is the separability of the effects of variation in weights and cluster sizes from the intraclass correlation. Measures of intraclass correlation vary widely across topic and segment size. None of the formulae have a strong theoretic basis. When the average segment size is small and intraclass correlation is moderate, DE gives nearly the same results as DE_2. Gabler, Haeder and Lahiri (1999) give circumstances under which DE_4 is conservative. They assume that intraclass correlation is constant across all clusters. This does not seem like a very good assumption to us. Hansen, Hurwitz and Madow (1953) argued in Section 4 of Chapter 8 of Volume 2 that variation in cluster size should increase the variance. Verma's DE_3 is interesting. He credits other statisticians with the suggestion but does not mention them by name or give any empirical or theoretical justification.

Table 1 shows the growth in size for the 30 segments in the NSPY sample that grew the most. On the NSPY listing of 3540 segments, the largest observed growth factor was 31.6 where a segment with just 64 dwelling units in it had 2019 dwelling units by late 1999. Segments with such extraordinary growth rates are rare. With a sample of 3540 segments, the next largest growth factor was 8.5, and there were only 13 segments with growth factors of 4 or more. Despite their rarity, these segments do cause major problems.

Table 1. Growth factors for 30 segments with strongest growth

Rank	HU90	HU99	Growth factor	Rank	HU90	HU99	Growth factor
1	64	2019	31.5	16	1218	4479	3.7
2	65	552	8.5	17	89	322	3.6
3	81	678	8.4	18	380	1360	3.6
4	80	537	6.7	19	112	385	3.4
5	63	398	6.3	20	81	278	3.4
6	100	606	6.1	21	73	245	3.4
7	237	1271	5.4	22	109	361	3.3
8	1156	5450	4.7	23	105	345	3.3
9	191	869	4.5	24	112	366	3.3
10	231	1019	4.4	25	290	946	3.3
11	85	362	4.3	26	124	398	3.2
12	61	253	4.1	27	87	278	3.2
13	81	327	4.0	28	296	942	3.2
14	96	369	3.8	29	383	1208	3.2
15	460	1698	3.7	30	195	605	3.1

If, at one extreme, we never truncated segment sample sizes, the relative variance in weights would be zero and the relative variance in segment sizes 0.39. If at the other extreme, we truncated segment sample sizes at the expected size (i.e., g=1), the relative variance in weights would be 0.34, and the relative variance in segment sizes would be 0.02. So at either extreme, the design effect would appear to be well above 1.3. For intermediate values of g such as g=4, the sum of the two relative variances is smaller, leading us to question the accuracy of the approximation. For g=4, we have a relative variance in weights of 0.05 and a relative variance in segment sizes of 0.13, suggesting a total design effect of 1.18. We suspect that it is possible for the two sources of variation to interact together in ways that makes the total design effect larger. We thus chose 1.3 as a likely design effect for an area sample that chose not to use permit sampling.

Note that a design effect of 1.3 can be extremely important for a large survey or several years of a recurring survey. For a recurring survey with a 10-year data collection budget of $100 million for example, eliminating such a design effect would be worth $23 million. (The sample size and thus the variable budget could be adjusted by a factor of 1/1.3.) This is clearly enough potential cost savings to be worth considerable complication in the sampling procedures. On the other hand, if the total budget for data collection in a one-time survey is $2 million, then the potential savings are only on the order of $500,000. Moreover, unless listing is done near the time of the Census, there must be an allowance for the cost of screening out new construction in the area segments.

3. Costs of Screening out New Construction in Area Segments

In this section, we express the cost of screening out new construction in terms of its relationship to the cost of regular screening and interviewing. The screening question itself is quite simple. Something along the lines of when was this house/apartment building built? Usually, the question is not asked for mobile homes since building permits are not generally required for mobile home placements. Of course, the importance of the extra screening in the overall budget will depend on the length of the interview and whether there are any special procedures that need to be done on sample persons such as testing of hair or urine or conducting physical examinations as in the National Health and Nutrition Examination Survey (NHANES). Let R be the ratio of the cost of conducting an extended interview on a person (including the screening cost for that particular unit) to the cost of just screening the unit and then discarding it.

It has been Westat's experience that for a one hour interview, screening costs about one third as much as interviewing. The proportion of housing that was built in the 90s and would thus need to be screened out of the area sample was about 12 percent or 1.1 percent per year. So a survey selected t years after the Census can expect a penalty of

$$\frac{(1.011)^t - 1}{R}.$$

For example, if R=3 and t=10, then the cost of screening will lead to a cost increase of 3.9 percent. This would then reduce the cost savings due to using permit sampling from 23 percent to 20 percent. Larger values of R imply more savings and smaller values imply reduced savings, but even if R=1, there is still a net savings of about 14 percent from using permit sampling.

4. Costs of BPO Selection and Listing

The work to select a sample of building permits is complex and costly. The first step is to select a sample of building permit offices. Within these, a sample of segments must be selected. After selection of the segments, it is necessary for a lister to visit all the multi-unit building sites within the selected permits and to list all the housing units that are found there. After listing of housing units within sample segments, the final sample of dwelling units may be selected. Much of the complexity is due to the unevenness of building activity across the country and across time. Also complicating the process is the rich variety of residential structures that are erected across the nation, the lack of uniformity in permitting practices, and the uncertainties in the construction business itself that can lead to wide variation in the time gap between the granting of a permit and the occupancy of the structure.

The Census Bureau smoothes some of the variation out in the practice of collecting permit data. Each month, staff from the Bureau collect counts of permit activity in all the permit offices with high activity levels across the nation. Once a year, the Bureau collects counts permit activity form the less active office. These counts are by type of structure. Counts are obtained both of permits and of permitted units. These counts can be quite different when permits are issued for apartment buildings. Some offices count townhouses as multi-unit structures, but these are reclassified by the Bureau as single-unit structures. The first stage in permit sampling is to restrict the sample of permit offices to those within sample PSUs. Since the jurisdictional boundaries of an office need not respect PSU boundaries, there are some decisions to be made of which PSU an office should be associated with.

The second stage of sampling is to select a sample of permit offices. When sampling for multiple waves, panels, or years of a survey, it is particularly cost effective to select a sample of permit offices before selecting the sample of permits. The reason for this is that a lister will need to gain access to every sample permit office. The procedure of finding the office, introducing oneself, gaining access to the records, and finally abstracting the data can be time consuming. Most often, it is welcome when this cost can be amortized across several waves of data or even across several different surveys. Some offices, however, have such low levels of permit activity that they need to be collapsed with other offices in order to provide enough sample for all the waves or surveys that are being jointly sampled. (For NPSY, we also excluded permits issued by offices with fewer than 30 permits issued over a 9-year period. This resulted in a small coverage loss of 0.7 percent of permitted new construction or 0.08 percent of all housing. In NSPY, we succeeded in getting cooperation from every one of the sample offices although this meant a lot of work in some cases.)

The third stage of sampling is the segment. Since most permit offices used to file their permits by month of issuance, the natural sampling unit within offices has been the office-month since the development of the technique. In addition to reducing sampling costs, this also reduces interview costs since units with permits granted in the same month by the same office are often in a small number of developments. We found that most offices became automated within the past decade, so the filing structure is less important than it used to be, but it is probably still the case that electronic archives are organized with some chronological structure. For the monthly offices, the monthly reports allow one to easily assign a measure of size to each office-month. For the annual offices, it is natural to take the entire year as a segment although another option is to impute a measure of size to each month within the year. Most surveys require a minimum size for each segment. To achieve this in monthly offices with lower than usual activity, it is often necessary to collapse consecutive months. (In NSPY, we succeeded in getting a list of permits for 94.2 percent of the segments selected. The failures were mostly in offices that had switched from paper record systems to electronic record systems and were unable or unwilling to find the old paper records.)

The fourth stage of sampling is the permit. Permits should be selected with probability proportionate to number of units contained within them. Rather complex sampling procedures are needed to ensure that the collection of sample permits will provide enough dwelling units. At the same time, the procedures must be capable of being executed by listers while they are in the permit offices. It is too costly to have them list all the permits within the sample segments. After selection of sample permits, the listers visit the building sites for multi-unit permit to list all the individual apartments.

The fifth and final stage of sample selection is to select the final sample of dwelling units; i.e., houses and apartments. Once the sample of dwelling units is selected, the interviewing procedures are quite similar to those from area segments. The only difference is that interviewers must be trained on how to track permits for structures that were not built yet at the time that the permit lister collected the permits or which the permit lister was unable to find. (On NSPY, we obtained a response rate to the first brief screening question of 92.0 percent on the permit sample, only a little worse than the comparable rate of 95.4 percent for the area sample. We speculate that most of the difference is due to problems in finding sample permits.)

The total costs for statisticians and programmers to conduct the five stages of sampling are significant. However, the most expensive operation is to send the listers to the BPOs. A significant part of this cost is the hiring and training of the listers as well as the preparation of training materials. Also, there are some data management costs for the lists of DUs.

5. Biases

Various other papers in the past have pointed out some of the biases associated with permit sampling. According to Census Bureau construction statistics on housing starts, an average of 1.3 million housing units were built per year between 1990 and 1998. Only about 90,000 of these annual starts were outside the jurisdiction of any permit office up through 1994. In 1995, the Bureau expanded the census of building permit offices from 17,000 BPOs to 19,000. Since then, only about 48,000 housing units have been built annually outside of the jurisdiction of any BPO. This figure is estimated by the Census Bureau from a national sample of segments that listers periodically resurvey, looking for units built since the last visit. Little is known about how occupants of housing built outside of permit issuing areas are different from the rest of the country other than that almost all the non-permitted new construction is single family and in the South and Midwest. (For NSPY, these units were not given any chance of selection, thereby increasing undercoverage by about 0.5 percentage points.) The Census Bureau avoids this bias by using a different screening question in the area sample for area segments outside the jurisdiction of any permit office. In these segments, the Bureau does not screen on year built. We considered this procedure for NSPY, but in order to execute it, it would be necessary to know which Census blocks are within BPO boundaries and which are without. The Census Bureau builds such a list after every decennial census and maintains it during the decade, but it has not yet shared the list with outside contractors. Adding the loss from non-permitted starts on to the exclusion of offices with very low activity levels, the total coverage loss for NSPY was about 0.6 percent of housing.

Finally, there are inherent difficulties in ascertaining the age of buildings. Often the current occupants do not have good information on the age of the building, even when the building is less than 10 years old. For coverage to be exact, one would need to know when the structure was permitted. When there is a long gap between permit issuance and building construction, then there is undercoverage. When there is a shorter than expected gap, then there is overcoverage. Little is known about the impact of such coverage errors, but one should note that small errors of under-and over-coverage also exist in area listing due to uncertainties about segment boundaries and uncertainties about the residential nature of buildings.

6. Cost Comparisons

As discussed in Section 2, deterioration causes a design effect penalty of about 30 percent. So if permit sampling were free, it would be possible to cut costs by about 23.1 percent. However, permit sampling has two distinct costs. One is the cost to select the sample. The other is the cost to screen our new construction from the area sample. In a survey where the interview costs outweigh screening costs, this would be inconsequential. In a survey with large screening costs, it can be important. In the context of a survey with very large screening costs, the screening new construction out of the area sample boosts cost by about 6.5 percent, so the net cost reduction with permit sampling is about 18 percent.

The initial cost of the NSPY BPO sample was about the same as screening 27,000 dwelling units and then conducting 5600 interviews at 2300 homes. So if the sample size is on the order of 150,000 DUs, then the cost of the BPO sampling can be recovered. In surveys where there is no screening, just a household interview of moderate length (say an hour), the cost per complete is more like three times as high, so a sample size of 50,000 dwelling units would be sufficient to offset the fixed BPO selection costs. If the sample sizes are larger yet, then of course, the benefits strengthen. On the other hand, for smaller surveys, there do not appear to be sufficient savings to warrant the high fixed costs of permit sampling. In the case of NSPY, 270,000 DUs were to be screened, so the design is cost effective.

For recurring surveys such as the National Health Interview Survey, these costs are easier to justify since the 10-year sample size is on the order of 600,000 interviewed dwelling units; however, the fixed costs rise with recurring surveys since the sample permit offices must be visited annually to select recently issued permits.

7. References

Bell, B., Mohadjer, L., Montaquila, J., and Rizzo, L. (1999). Creating a frame of newly constructed units for household surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, p. 306-310.

Gabler, S., Haeder, S., and Lahiri, P. (1999). A model based justification of Kish's formula for design effects for weighting and clustering. *Survey Methodology*, **25**, p. 105-106.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). Sample Survey Methods and Theory. New York: John Wiley & Sons.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons.

Kish, L. (1987). Weighting in Deft^2. *The Survey Statistician*, June 1987.

Montaquila, J.M., Bell, B., Mohadjer, L., and Rizzo, L. (1999). A methodology for sampling households late in a decade. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, p. 311-315.

President's Committee to Appraise Employment and Unemployment Statistics. *Measuring Employment and Unemployment*. U.S. Government Printing Office, Washington, D.C., 1962.

U.S. Bureau of the Census. *The Current Population Survey – A Report on Methodology*. Technical paper No. 7. U.S. Government Printing Office, Washington, D.C., 1963.

U.S. Bureau of the Census. *The Current Population Survey – Design and Methodology*. Technical paper No. 40. U.S. Government Printing Office, Washington, D.C., 1977.

Verma, Vijay. (1993). Sampling Errors in Household Surveys. New York: United Nations Department for Economic and Social Information and Policy Analysis, Statistician Division, National Household Survey Capability Programme.

Waksberg, J. (1998). The Hansen Era: Statistical Research and Its Implementation at the Census Bureau, 1940-1970. *Journal of Official Statistics*, **14**, p. 119-136.

Waksberg, J. and Northrup, D.R. (1985). *Integration of Sample Design for the National Survey of Family Growth, Cycle IV, with the National Health Interview Survey*. National Center for Health Statistics. Vital Health Stat 2(96).