A METHODOLOGY FOR SAMPLING HOUSEHOLDS LATE IN A DECADE

Jill M. Montaquila, Bridgett Bell, Leyla Mohadjer, and Lou Rizzo, Westat
Jill M. Montaquila, Westat, 1650 Research Boulevard, Rockville, Maryland 20850

Key Words: New construction; Segment; Area sample; Measure of size

1. Introduction

Area sample surveys typically use information from the decennial census to select nationally representative sets of areas. In area sample surveys, the first stage of selection is the primary sampling unit (PSU), which is usually a county or groups of counties. Often, the second stage of selection is a smaller geographic area, referred to as a segment. Segments usually comprise blocks or groups of blocks. Further stages of selection may be done within segments.

Segments are ordinarily selected with probability proportionate to a measure of size. This is done in order to yield self-weighting estimates. If the measure of size is highly correlated with the item(s) being measured, then selecting segments with probability proportionate to a measure of size may also reduce the variance of estimates for that item. The measure of size may be total population, total number of households, a weighted population (weighted to give more weight to population subdomains of interest), or some similar measure. Population and housing estimates for sub-county areas such as groups of blocks are available only from the decennial census. One of the challenges of conducting area sample surveys late in a decade is that census data are outdated.

Early in a decade, it is reasonable to assume that the census data provide a good reflection of the segment measure of size. However, as the decade wears on, the decennial census data become less reliable due to growth from new construction and population declines due to demolition or housing vacancies. Therefore, late in a decade, segment measures of size based on decennial census data may be quite inaccurate.

Probability samples may be classified into two groups: Fixed-rate and fixed-size samples. With fixed-rate samples, a fixed, predetermined sampling rate is used and the sample size may vary. With fixed-size samples, a fixed sample size is specified and the sampling rate may vary.

Inaccurate segment measures of size will create problems for both of these types of probability samples. With fixed-size samples, large variations in segment sizes result in large variations in probabilities of selection of the ultimate sampling units, which in turn reduces the precision of estimates based on the sample. With fixed-rate samples, large variations in segment sizes results in large variations in the numbers of ultimate sampling units selected in each segment. This may result in sample sizes that are considerably different from anticipated sample sizes; it may also have operational and cost implications.

2. Using a Building Permit Frame to Sample New Construction

One way of reducing variations in segment sizes late in a decade is to control for new construction, the largest source of variations. A method that has been used to control for new construction in several area sample surveys conducted by Westat is building permit sampling.

With building permit sampling, a frame of building permit data is constructed. This frame is used to form segments comprising newly constructed units built since the last decennial census. The regular area segments are assigned measures of size based on the last decennial census, and during screening, newly constructed units are excluded from the regular area segments. The new construction segments are formed using data obtained by the Census Bureau for the Building Permits Survey (BPS), a monthly survey of building permit offices. The survey obtains information on the number of residential building permits issued in each reporting period. For large permit-issuing offices, the reporting period is typically a month, while for the smaller permit-issuing offices, the reporting period is typically a year. A complete segment frame that contains both the regular area segments and the new construction segments is constructed, and segments are sampled from this frame.

In regular area segments, dwelling units (DUs) are listed by canvassing the area. In new construction segments, DUs are listed by obtaining building permit records and listing the units corresponding to the permits issued during the selected time period. The two listings are then combined in order to create the frame of DUs. DUs are then sampled, typically by sorting the DU frame geographically (or according to permit number, for the new construction units) and then selecting a systematic sample. For more details about the listing, frame construction, and sampling processes, see Bell et al. (1999).

While the building permit information obtained by the BPS is useful for construction of a frame for

sampling persons residing in new construction, there are a few shortfalls. Building permits are not required for the placement of mobile homes. The permit files contain counts of the numbers of units for which building permits were authorized; a small fraction of units for which permits are authorized are never built. An additional concern relates to obtaining accurate measures of size based on the data from the BPS, since some of the data are imputed. Additionally, areas and units that do not require building permits for new construction are not included. According to the Census Bureau (U.S. Bureau of the Census 1994), in 1994, 5 percent of the U.S. population lived in jurisdictions that do not require building permits. This percentage of the population is heavily concentrated in the South Central and Great Plains areas: 26 percent of the population in the East South Central Census division, 17 percent of the population in the West South Central Census division, and 11 percent of the population in the West North Central Census division are contained in these 'non-permit' jurisdictions, with less than 2 percent of the population of the remaining Census divisions in non-permit jurisdictions. However, for the nation as a whole, less than 5 percent of all privately owned housing units are constructed in areas that do not require building permits.

3. Features of Building Permit Sampling

Permit sampling is probably the most cost-efficient method of updating an area sample frame. There are, however, some features and limitations related to this approach that must be taken into account when designing an area sample that includes permit sampling. In this section, we describe some advantages and disadvantages of permit sampling.

The primary advantage of building permit sampling is that it facilitates control over variations in segment sizes. As mentioned above, this allows for improved precision of the estimates. It also allows for more optimal, balanced caseloads and tighter cost controls.

There are a few disadvantages of building permit sampling that should be noted. First, the success of this approach depends upon being able to obtain accurate permit information. In our experience, we have found the permit files obtained from the Census Bureau to be quite accurate. If new construction segments are selected, the listers must obtain specific permit information from local building permit offices. Their success in doing so depends in part on the willingness of the permit office officials to cooperate. Another concern is that the permit files and records provide information for permits issued. Although a permit is issued, the unit may not have been built.

An operational concern with building permit sampling is the screening out of new construction in the regular area segments. In areas with a particularly high rate of growth, a large proportion of units in regular area segments may be screened out. Additionally, new construction segments are not clustered by geography. They are clustered by the time period in which the permits were issued. Thus, a considerable amount of effort may be expended in traveling to and locating the new construction units. A related concern is that the boundary between the regular area sample and the new construction sample is not perfectly defined. The screening out of the new construction in regular segments relies on the respondents' knowledge of when their residential buildings were constructed.

An analysis of data from a survey conducted by Westat in the late 1980's that used the permit sampling method revealed differences between the 1990 census data on the estimated number of persons residing in buildings constructed between 1980 and 1990 and the survey estimate of total population residing in newly constructed units. Some of the discrepancy was expected to be attributable to differences in the method of data collection. The census estimate relies heavily on the respondents' knowledge of when their residential building was constructed. This estimate is based on the sample data and includes imputed values for missing data. Furthermore, estimates from the data reported by permit offices exclude about 5 percent of housing units that are built in the U.S. Persons residing in units that do not require building permits are likely to be somewhat different from those residing in buildings with authorized permits. Sample weighting adjustment methods are usually used to compensate for such differences.

In the National Survey of Parents and Youth (NSPY), we sampled a total of 130 building permit offices from the 90 primary sampling units to provide a nationally representative sample of households in housing constructed after 1990. (For details, see Bell et al., 1999.) From these offices, a total of 516 building permit segments were randomly selected, with each segment corresponding to all building permits issued within a particular time interval (e.g., February 1996 through April 1996) for that office. We were generally successful in getting the detailed information necessary from the building permit offices for these sampled segments, but in some cases the necessary data was found to be inaccessible (as it was destroyed or too difficult to access), or the office was not cooperative for a variety of reasons. The table below presents our success rate: roughly 6 percent of the segments were not recoverable; these will contribute to the NSPY nonresponse rate (in addition to household and person level nonresponse). Multiplying the 6 percent loss rate by the overall building permit share of the sample

(roughly 10 percent) gives 0.6 percent, the contribution of the 30 lost segments to overall NSPY nonresponse.

Building permit segments	Count	Percent of total
Completed	486	94.2
Inaccessible	13	2.5
Other	17	3.3
Total	516	100.0

4. Alternative Approaches

Permit sampling works well when the amount of new construction is moderate, and when the data collection period for the survey is not too restricted. Recently, Westat has considered alternatives to building permit sampling. Two alternatives under consideration include using a two-phase sample of segments, where the first phase is used to obtain updated segment measures of size; and contacting local officials to obtain information about growth that is used to update the segment measures of size.

Double sampling (or two phase sampling) is used extensively for stratification, and regression or ratio estimation. One of the uses of double sampling is to update a sampling frame when the sample is to be selected with respect to a measure of size (MOS) but a reliable estimate of the MOS is not available. With this approach, a larger sample of units (in this case, area segments) is selected. Updated values of the MOS are then collected for this larger sample (also referred to as the first phase sample). The final sample of units (segments) is selected from the first sample using the updated MOS.

There are of course costs associated with the double sampling procedure. (There are costs associated with the permit sampling also, but most likely, the costs associated with double sampling are somewhat higher.) As mentioned above, the updated MOS obtained from the first sample is used to improve the final sample of segments. This approach is beneficial if the advantages or improvements in the selected sample more than compensate for the cost of conducting the first phase sample. More specifically, double sampling is practical only if the per element cost of obtaining the first phase measurement is much smaller by a large factor than the per element cost of the principle measurement.

A hybrid of these two approaches has been tested for the Community Longitudinal Survey of Parents and Youth (CLSPY). Four sites, having varying levels of new construction, were selected for this test. In this test, in PSUs with large percentages of new construction, local officials (typically officials in local planning offices) were contacted and asked to identify Census blocks having a large amount of new construction. Listers cruised these blocks to obtain updated dwelling unit counts, and these updated counts were used to revise the measures of size for these blocks. Two of the four sites were deemed large enough to warrant contacting the local planning offices to identify the high-growth blocks. In this test, mixed results were obtained from contacting the local planning offices. In one PSU, fairly comprehensive information was obtained about all of the new developments in the 1990s; in the other PSU, the information provided was mainly about very recent development, and very little information was provided about new construction during the earlier part of the decade.

5. Implementing the Double Sampling Approach

In this section, we describe the basic goal of the double sampling (or two-phase) approach to segment selection and outline its implementation. The primary objective of the double sampling approach is to obtain a segment measure of size that is as accurate as possible, by first using information from the building permit files and later using direct counts obtained for the Phase 1 segments.

5.1 The Phase 1 Sample

Let M_i denote the measure of size for segment i based on data from the most recent decennial census. (For ease of representation, the subscript indexing the PSU has been dropped for this discussion.)

Using building permit data, we can obtain estimates of "growth" in places having permit-issuing offices. Let U_p denote the number of units for which building permits were issued since the most recent census in place p. Note that U_p is available only for places with permit offices and is obtained directly from the building permit files compiled by the Building Permits Survey. The number of persons residing in newly constructed units in place p may be estimated by $T_p^{[u]} = 2.6U_p$. The estimate of the place-level "growth" ratio is then given by

$$g_p = \frac{T_p^{[o]} + T_p^{[u]}}{T_p^{[o]}},$$

where $T_p^{[o]}$ is the total population for place p from the most recent decennial census.

For the first phase of segment selection, the measure of size for each segment is adjusted for the place-level growth (which is used as a proxy for segment-level population changes). That is, the measure of size for segment i in the Phase 1 segment selection is M_i', where

$$M_i' = M_i g_p, i \in p .$$

Because M_i' is an estimate of the true segment measure of size that is based on place-level data, it is subject to error. The second phase of segment selection (described below) will correct for errors in M_i'. However, in order to ensure that target overall sampling rates or sample sizes can be attained, it is necessary to select a Phase 1 sample of segments that is somewhat larger than the ultimate segment sample size and will enable within-segment target sampling rates or sample sizes to be attained.

5.2 The Phase 2 Sample

Typically, segments are much smaller than places. Therefore, the estimated place-level growth may differ considerably from the true change in the size of the segment since the last decennial census. In order to obtain more accurate estimates of the true change in the size of the segment, a new "counting" procedure is used for the Phase 1 segments. "Counters" (experienced listers) travel to the Phase 1 segments and count the number of dwelling units in each of the segments. These counts are used to compute measures of size for the Phase 2 segment selection such that the overall probabilities of selection of the segments are accurate.

Let U_i' denote the number of DUs found by counters when counting Phase 1 segment i. The change in the size of Phase 1 segment i is estimated based on U_i' as follows:

$$g_i' = \frac{U_i'}{U_i^{[0]}},$$

where $U_i^{[0]}$ is the number of DUs in segment i at the time of the most recent decennial census.

The measure of size for Phase 2 selection is

$$M_i^{[2]} = \frac{M_i g_i'}{M_i'} = \frac{g_i'}{g_p}, \; i \in p$$

Let U denote the set of all segments on the frame for Phase 1 selection, and let S_1 denote the set of segments selected in Phase 1. The overall probability of selection of segment i (conditional on the sampled PSUs) is

$$P_i = \frac{k_1 M_i'}{\sum_{i \in U} M_i'} \cdot \frac{k_2 M_i^{[2]}}{\sum_{i \in S_1} M_i^{[2]}} = \frac{k_1 M_i'}{\sum_{i \in U} M_i'} \cdot \frac{k_2 M_i g_i' / M_i'}{\sum_{i \in S_1} M_i^{[2]}} = \frac{k_1 k_2 M_i g_i'}{\sum_{i \in U} M_i' \sum_{i \in S_1} M_i^{[2]}},$$

where k_1 and k_2 are the numbers of segments selected in the Phase 1 and Phase 2 samples, respectively.

5.3 Other Considerations

The main advantages of a double sampling scheme are:

1. All segments become regular segments. There are no new construction segments. This eliminates the need to contact the permit offices and reduces the amount of traveling needed for interviewers to locate the selected dwelling units (DUs).
2. The screening question about whether the household was built after 1990 is eliminated. Thus, interviewers' time is not spent on trying to contact and get cooperation of DUs that eventually will become ineligible because the DU was built after 1990.

Ideally, a very large sample of segments should be selected in the first phase so that any unusual segment growth can be incorporated into the final segment sample. In simple words, for example, if a segment has grown to about five times the 1990 size, then a sample of about five times larger than the final segment sample is needed to appropriately account for such growth without having a direct effect on the precision of the survey estimates. Depending on the resources available to a survey, it may be practical to select a first phase sample of about 5 times or even 10 times larger than the final segment sample. However, for any given survey, the cost of selecting a first phase sample may become prohibitive if the size of the first phase sample exceeds some threshold.

Moreover, the Phase 1 sample is a function of how variable the rate of growth is between segments in a PSU. If the rate of growth is consistent among all segments then the Phase 1 sample will be equal to the final segment sample. However, if the rate of growth is very variable, then we will need to select many more segments from the highly developed areas within the PSU to make up for their increase in size. We are currently evaluating the use of within PSU permit information (available from Census) to stratify the Phase 1 segments (within PSUs) and to select the Phase 1 sample as a function of the growth within various areas in a PSU.

6. Conclusions

Because persons residing in newly constructed units may be systematically different from those residing in older units, adequate coverage of such persons is often essential. For area sample surveys, one method that is frequently used to cover persons residing in new construction is building permit sampling. In Section 3 we discussed the advantages and limitations of permit sampling.

We have presented an alternative to permit sampling — a double sampling methodology that uses information collected for the first phase sample of segments to update the measure of size for selecting the second phase sample. We believe this method shows promise. We have conducted a small-scale test of this method in conjunction with one area sample survey, and we plan to implement this method for another area sample survey beginning in 2000.

7. References

Bell, B., Mohadjer, L., Montaquila, J., and Rizzo, L. (1999). Creating a frame of newly constructed units for household surveys. To appear in *Proceedings of the Survey Research Methods Section.* Washington, DC: American Statistical Association.

U.S. Bureau of the Census (1994). *Current Construction Reports — Housing Units Authorized by Building Permits: Annual 1994*, Publication C-40/94-A, U.S. Government Printing Office, Washington, DC.