A MODIFIED HOUSING-UNIT METHOD FOR SMALL-AREA POPULATION ESTIMATION

Norfleet W. Rives, Jr., University of Delaware

1. INTRODUCTION

The College of Urban Affairs and Public Policy at the University of Delaware maintains an extended survey capability to support various planning and evaluation projects conducted by State and local public agencies. A central element of this survey capability is the New Castle County household address file. The University is located in New Castle County, the largest county in the State. The estimated 1975 population of the county is 395,000. The household population is distributed among approximately 130,000 housing units. The county household address file contains information on physical location for each house, apartment, or other structure occupied or intended for occupancy as separate living quarters.

The ability of the household address file to support different programs of survey research in New Castle County is well documented in numerous reports published by the College of Urban Affairs and Public Policy. These reports typically summarize the findings of household surveys conducted on a large scale, but, more recently, and with further improvement in the accuracy and completeness of the basic file, several projects on a smaller scale have been undertaken. One project is concerned with the estimation of local population.

The procedure involved in using the county address file to estimate population is based on the same principle as the housing-unit method for estimating the postcensal population of urban areas (Shryock and Siegel, 1971). The housing-unit method produces a figure for total resident population by combining information on (1) the stock of housing units, (2) average household size, (3) the vacancy rate, and (4) the population in group quarters. The estimating equation is:

$$P = bH(1-w)/(1-g), \qquad (1)$$

where: P = total resident population,

H = stock of housing units,

b = average household size,

w = vacancy rate, and

g = proportion of population in group quarters.

The usual method for estimating H is given by the expression:

$$H = H_c + H_a - H_d, \qquad (2)$$

where: H_c = stock of housing units in the most recent census,

H_a = housing units added to stock during the postcensal period, and

H_d = housing units demolished between census date and the estimation date.

Various sources of information may be used to obtain numerical values for b, w, and g.

The accuracy of the conventional housing-unit method clearly depends upon the accuracy with which the components of equation 1 can be estimated. The usual procedure in many areas is to derive estimates for b, w, and g from the most recent census. This procedure is simple and inexpensive, even though the dated usefulness of census information is well known. The usual procedure for estimating H is to update the census figure (H_c) with information collected locally on housing-unit completions and demolitions. According to several recent studies on the accuracy of different postcensal population estimation methods, the quality of completion and demolition certificates is subject to considerable variation at the local level, and this is a principle reason for the frequently poor performance of the housing-unit method (Morrison, 1971; Starsinic and Zitter, 1968).

The use of a household address file to estimate the current stock of housing units will certainly not overcome all of the problems associated with the use of census statistics in combination with unit completion and demolition certificates, but the former approach will not be any less effective, and, depending upon how the file is updated, it may actually be more effective. Segments of the New Castle County file, corresponding to specific geographic areas, are periodically re-field-listed in connection with particular survey projects, but this is not the primary method for continuous monitoring. The general file is routinely updated using (1) completion and demolition certificates, (2) public utility records on the connection and disconnection of electric meters, and (3) regular reports from survey personnel on the actual status of housing units encountered in the course of conducting interviews. This particular combination of information has been found to facilitate the maintenance of the address file, and, therefore, it should facilitate the estimation of the total number of county housing units on a given date.

If one can assume that a household address file provides a reasonably accurate and complete specification of a particular stock of housing units, then one is in the rather fortunate position of being able to construct a useful population estimate with a relatively simple statistical procedure. Population estimates produced by this procedure are designated modified housing-unit estimates, because the procedure is mechanically similar to the conventional housing-unit method. The following section presents the theoretical foundations of the proposed method, and the final section illustrates its application to a particular city in the State of Delaware.

2. THEORETICAL FRAMEWORK

Let H be the total number of housing units in a perfectly specified household address file for a given area, and let X_i be the number of persons who usually reside in the ith unit; vacant units will have X_i equal to zero. Suppose now that one samples this file, selecting n household addresses at random. The sample may be either simple random or systematic, depending upon whether one is willing to assume that the ordered series $(X_1,\ldots,X_H)$ forms a random sequence. In practice, this will usually be a tolerable assumption, primarily because systematic samples are less expensive to select and routinely produce only slightly larger variance terms than those obtained from simple random samples (Kish, 1965).

Information on the number of persons who usually reside at each household address in the sample group may be obtained either by mail or by personal interview. The mail approach is less expensive, but more difficult to control, and it may produce a troublesome level of nonresponse, even though one can hardly expect respondents to become disgusted and impatient with a simple prepaid postcard containing a pleasant explanatory note, a single non-threatening question, and possibly some official emblem such as a university seal. The interview approach is more expensive, but it is less difficult to control and can usually be relied upon to produce the desired information with relatively few problems. If only a single question is involved, however, as in the present case, then an interview survey can hardly be justified, unless the interviewing process is virtually costless. Conversely, if the single question on household size can be appended to a larger survey which has the proper sample design and would have been conducted in any event, then the marginal cost of interviewing to obtain household size data for the sample group is relatively small.

When the household address file has been sampled and the household survey completed, one will be left with the statistical series $(X_1,\ldots,X_n)$. Average household size (b) can then be estimated from the equation:

$$b = (X_1+\ldots+X_{n-k})/(n-k) , \qquad (3)$$

where k is the number of vacant units in the sample. The estimated vacancy rate is given by the expression:

$$w = k/n . \qquad (4)$$

A point estimate of the total population in households (P_h) can now be obtained from the equation:

$$P_h = bH(1-w) . \qquad (5)$$

This estimate, combined with an independent estimate of the population in group quarters, can be used to produce a final point estimate of total resident population. If g is the proportion of the area population in group quarters, then total resident population (P) is given by the expression:

$$P = P_h/(1-g) . \qquad (6)$$

To provide for the direct estimation of P from the sample data, equation 6 can be rewritten in the form:

$$P = bH(1-w)/(1-g) . \qquad (7)$$

A useful source of information on the proportion g is the most recent census. If the local group-quarters population is small in relation to the total resident population, as is typically the case, then the census proportion can be safely applied in equation 7. In areas where the population in group quarters is proportionately large, this population will almost always have a primary source, such as an institution of higher education or a correctional facility. Information on the size of these populations can normally be obtained from the appropriate institutional officials, but, in the case of student populations, one should be careful to select only students who are area residents and not otherwise subject to the risk of having been included in the household survey. If the area under study contains a military installation, one should obtain a separate estimate of the total resident population of the installation from appropriate installation personnel. Information of this type is routinely available for official purposes, with the approval of the installation commander.

An important methodological issue in the estimation of average household size (b) and the vacancy rate (w) is the determination of sample size. The number of housing units surveyed will directly affect the cost of the project, the time required to complete it, and the statistical precision of the population estimate.

Assume that an estimate of the true vacancy rate (R) is desired accurate to within a fixed proportion E, with statistical precision given by the standard normal variate z. Under these circumstances, the minimum sample size (n_w) can be obtained from the expression:

$$n_w = \frac{Nz^2R(1-R)}{(N-1)E^2 + z^2R(1-R)} . \qquad (8)$$

This expression is the familiar equation for the determination of sample size when the sampling frame is finite (Lapin, 1975).

Equation 8 cannot be solved without a value for R, and this puts one in the rather awkward position of needing the value of a parameter to find the value. Actually, the value for R in equation 8 is an intermediate value in the calculation procedure, and one can afford to select a value in a rather casual manner. If some estimate of the vacancy rate cannot be obtained locally, then the most recent census may be a useful source of information. A less desirable solution is to use the maximum value of the product R(1-R) in equation 8. Since R is a fraction, the product R(1-R) will reach a maxima when R equals 0.5, but it is quite unlikely that a local vacancy rate for all structures combined would ever reach 50 percent.

The minimum sample size required to estimate average household size can be found using an

expression similar to equation 8. Let V(X) be the variance of household size, and assume that an estimate of true average household size is desired accurate to within F persons per unit, with statistical precision again given by the standard normal variate z. The sample size in question is the solution to the equation:

$$n_x = \frac{Nz^2 V(X)}{(N-1)F^2 + z^2 V(X)} . \qquad (9)$$

This equation, like equation 8, is the usual expression for the determination of sample size when the sampling frame is finite (Lapin, 1975).

The variance term V(X) in equation 9 should ideally be the population variance, but the actual parameter is almost always difficult to obtain. Census tabulations on household size could be used to construct a reasonably good approximation to V(X), but this approach will be subject to one rather important limitation: Since the most recent census in 1970, there has been a significant increase in the rate of formation of primary (single-person) households in the United States (Kobrin, 1976). This means that the variance of household size has almost certainly declined during the post-censal period. The extent of any such reduction will obviously depend upon the particular local population being studied. In general, the estimated variance of household size obtained from the 1970 Census should be considered a maximum value, and, indeed, one may even want to use a smaller figure in the calculation for sample size.

The minimum sample size required to estimate the vacancy rate with a given statistical precision will rarely equal the minimum sample size required to estimate average household size with the same precision. Since the household survey operates on a single sample, the estimates from equations 8 and 9 must be reconciled. The most defensible solution to this problem is to select the larger of the two sample sizes. This will insure a certain minimum level of statistical precision for one of the parameter estimates and a somewhat higher-than-desired level for the other. Increasing the number of housing units to be surveyed beyond some lower limit should not be a cause for concern, because the marginal cost of sampling is usually quite small, especially when the survey is conducted by mail.

3. EMPIRICAL APPLICATION

The modified housing-unit method to estimate local population was originally evaluated using census and survey statistics for the City of Newark, Delaware. The Newark household address file was the first component of the New Castle County file, with other areas being added to the system between 1968 and 1973. The procedure for updating the Newark file is well established, and the quality of the information used in this process has been subjected to rigorous testing on more than one occasion. The most recent estimate of the total resident population of Newark was prepared for the midyear date, July 1, 1975. The following discussion summarizes the estimation procedure.

The Newark household address file contained 6,509 housing unit locations on July 9, 1975, the day the survey sample to estimate local population was actually drawn. Since the Newark file had been updated only five weeks before this date, no minor adjustments were made for file specification errors. Under other circumstances, it might have been necessary, or at least advisable, to adjust the file for known biases.

Prior to selecting the sample, it was decided that the vacancy rate should be estimated to within 1 percent and average household size to within 0.1 persons per unit, both with 80 percent statistical significance. According to an analysis of census tabulations, the approximate variance of household size in Newark in 1970 was 1.72. To allow for the effect of an increase in the proportion of primary households on the variance of household size, the census figure was reduced to 1.50; this adjustment is clearly arbitrary. The 1970 Census vacancy rate for the Newark area was less than 2.5 percent, but a 1974 estimate prepared locally put the rate closer to 3.0 percent. After some deliberation, it was decided to accept the more recent figure.

The minimum number of housing units required to estimate the vacancy rate with the desired precision was computed using equation 8. The solution value for n_w was 445. The corresponding solution value for the minimum sample size necessary to estimate average household size with the desired precision, following equation 9, was 237. According to the decision rule previously established, the minimum sample size for the survey should be 445, since 445 is obviously larger than 237. The actual survey sample consisted of 500 housing units, however, to allow for the effect of survey nonresponse. This adjustment was purposely smaller than the usual adjustment, because the survey, which was conducted by mail, contained only a single, relatively insensitive question. The usual over-sampling rate for mail surveys conducted by the College of Urban Affairs and Public Policy is closer to 20 percent.

Each household receiving a survey postcard was asked to provide a simple count of the number of persons who usually resided in the unit on July 1, 1975. The survey period was set at four weeks from the date of mailing, and, at the end of this period, 462 responses had been obtained. The Newark Post Office indicated that 15 units were vacant, and, according to other sources, 2 units had been demolished. The remaining 21 units were simply classified as legitimate nonresponses, although 3 postcards were received several months after the survey had technically been completed. No attempt was made to identify nonrespondent units.

The survey data were processed manually to produce estimates for the vacancy rate (w) and average household size (b). The estimated value of b was 2.94, and the estimated value of w was 3.1 percent. When these estimates are substitued into equation 5, one obtains an estimated household population of 18,553.

The final estimate of the total resident population requires only an estimate of the population in group quarters. The principal source of the group-quarters population in Newark is the University of Delaware. According to University enrollment records, the number of students to be added to the resident population of Newark was approximately 6,800. This figure was compiled from data for the preceding academic year and the first session of summer school.

If 6,800 were added to the population in households, then the estimated total resident population of the City of Newark on July 1, 1975, would be 25,353. The 1970 Census population was 21,078. This would make the average annual growth rate during the postcensal period 3.5 percent, an entirely plausible figure for a small metropolitan community in New Castle County.

REFERENCES

Kish, L. SURVEY SAMPLING. New York: Wiley, 1965.

Kobrin, F. "The Fall of Household Size and the Rise of the Primary Individual in the United States." DEMOGRAPHY 13 (1976): 127-138.

Lapin, L. STATISTICS: MEANING AND METHOD. New York: Harcourt, Brace, Jovanovich, 1975.

Morrison, P. DEMOGRAPHIC INFORMATION FOR CITIES: A MANUAL FOR ESTIMATING AND PROJECTING LOCAL POPULATION CHARACTERISTICS. R-618-HUD. Santa Monica: Rand, 1971.

Shryock, H. and J. Siegel. THE METHODS AND MATERIALS OF DEMOGRAPHY. 2 Vols. Washington: U. S. Bureau of the Census, 1971.

Starsinic, D. and M. Zitter. "Accuracy of the Housing Unit Method in Preparing Population Estimates for Cities." DEMOGRAPHY 5 (1968): 475-484.