DEMOGRAPHY© Volume 17, Number 3 August 1980

SOME NEW TECHNIQUES FOR APPLYING THE HOUSING UNIT METHOD OF LOCAL POPULATION ESTIMATION

Stanley K. Smith
Department of Economics and Bureau of Economic and Business Research, University of Florida, Gainesville, Florida 32611

Bart B. Lewis
The Atlanta Regional Commission, Atlanta, Georgia 30303

Abstract—The housing unit method of population estimation is often characterized as being imprecise and having an upward bias. We believe that the method itself cannot properly be characterized by a particular level of precision or direction of bias. Only specific techniques of applying the method can have such characteristics. In this paper we discuss several new techniques we have developed for estimating households and the average number of persons per household. Estimates produced by these techniques are compared to estimates produced by several other techniques. Special census results from Florida provide preliminary evidence that the new techniques produce more precise, less biased estimates than the other techniques.

INTRODUCTION

The housing unit method of population estimation is one in which population estimates are derived from estimates of occupied housing units and is the most commonly used method for making substate population estimates in the United States. A recent survey conducted by the Bureau of the Census showed that more than three-fourths of all agencies making substate population estimates use some form of the housing unit method (U.S. Bureau of the Census, 1978a, p. 7). Indeed, data limitations often make the housing unit method virtually the only method available to state and local demographers making subcounty population estimates.

The housing unit method is conceptually clear and theoretically sound. The population of any given geographic area is identically equal to the number of occupied housing units (households) times the average number of persons per household (PPH), plus the number of persons living in group quarters (e.g., college dormitories, nursing homes, penal institutions, military barracks):

$$P_t = (H_t \cdot PPH_t) + GQ_t \qquad (1)$$

Where:

P_t = total population at time t
H_t = occupied housing units at time t
PPH_t = average number of persons per household at time t
GQ_t = group quarters at time t

This is an identity not an estimate. If these three components were known exactly, the exact total population would also be known. As pointed out by Lowe, Pittenger and Walker (1977, p. 1), any failure to obtain exact population estimates from the housing unit method must be due to inaccurate data or imperfect application techniques, not to the method itself.

The basic problem, of course, is that these three components are never known exactly. They must be estimated from other data sources. Many techniques can be used to estimate each component; each technique has its own characteristics with respect to precision and bias. Population estimates based on the housing unit method can be very good or very bad, depending on the characteristics of the specific techniques and data employed.

A number of techniques are available for estimating each component but only a few have been thoroughly tested. By far the most widely cited test of housing unit techniques was performed by Starsinic and Zitter (1968). They evaluated several techniques for estimating households and persons per household by comparing estimates with the results of special censuses conducted in 47 cities with population greater than 50,000. Although this study was simply a test of several specific techniques, it has mistakenly come to be accepted as a test of the method itself. A number of Starsinic and Zitter's findings with respect to precision and bias have passed virtually unchallenged into general acceptance as characteristics of the housing unit method. This is unfortunate. The results reported by Starsinic and Zitter reflect characteristics of the particular techniques they employed, not of the housing unit method itself. While the Starsinic and Zitter techniques are certainly legitimate, they are by no means the only techniques that can be used for making housing unit population estimates.

In this paper we discuss some new techniques we have developed during several years of making local population estimates for the State of Florida. Special census results from 22 places in Florida are compared with estimates derived from the new techniques and estimates derived from the Starsinic and Zitter techniques. These comparisons provide preliminary evidence that the new techniques produce more precise, less biased population estimates.

METHODOLOGY

Equation (1) defines the population of any geographic area as the number of households times the average number of persons per household, plus the number of persons living in group quarters. In this section we describe our techniques for estimating these three components, the advantages and disadvantages of each and how they differ from the Starsinic and Zitter techniques.

Number of Households

Starsinic and Zitter evaluated two different techniques for estimating households, the first based on building permits. They estimated current housing units by adding residential units permitted since the most recent census to units enumerated in that census. They assumed that three months were required for permits to become finished units. The housing stock was adjusted downward to account for demolitions. Current households were estimated by applying the occupancy rates observed in the most recent census to estimates of current units:

$$H_t = (HU_c + BP_t - D_t) \cdot OCC_c \quad (2)$$

Where:

HU_c = housing units enumerated in most recent census

BP_t = units permitted between most recent census and time t (3 month lag)

D_t = units reported demolished between most recent census and time t

OCC_c = overall occupancy rate observed in most recent census

One problem with this approach is the failure to disaggregate by type of housing unit. Single family, multifamily and mobile home units have considerably different time lags between the date building permits are issued and the date units are ready for occupancy. Surveys conducted in Florida have shown three to five month time lags are most common for single family units and ten to eighteen month lags for multifamily units. Mobile home units are typically ready for occupancy in the month the permit is issued. Starsinic and Zitter's use of a three month time lag for all unit types is undoubtedly one reason their estimates of households showed an upward bias.

Vacancy rates also differ markedly by type of unit. Single family rates are generally much lower than multifamily rates. In the second quarter of 1979, for example, U.S. vacancy rates were nearly 6 percent for multifamily units and just over 1 percent for single family units (U.S. Bureau of the Census, 1979a, pp. 4–5). The use of a single overall vacancy rate could result in a large error for an area in which the mix of types of housing units has changed considerably since the most recent census.

When housing units are differentiated by type of unit, the number of households can be estimated as:

$$H_t = \sum_i (HU_{ic} + BP_{it} - D_{it}) \cdot OCC_{ic} \quad (3)$$

Where:

HU_{ic} = housing units of type i enumerated in most recent census

BP_{it} = housing units of type i permitted since most recent census (4 month lag for single family, 12 for multifamily, 0 for mobile home)

D_{it} = units of type i reported demolished since most recent census

OCC_{ic} = occupancy rate for housing type i in most recent census

i = single family, multifamily, mobile home units

Compiling building permit data by type of unit is simple but time-consuming. The data are available for both permits and demolitions for all areas which file Form C-404 with the U.S. Department of Commerce. It is estimated that approximately 90 percent of all new housing units in the United States are built in areas requiring building permits (Siskind, 1980, p. 1). In Florida this proportion is even higher. Places that do not issue building permits are generally small towns and sparsely settled unincorporated areas. For the places that continuously issue building permit data, the quality of the data is generally quite good except for mobile homes. Problems with mobile home data include the issuance of permits for spaces in mobile home parks rather than for mobile homes themselves, double-counting for changes in ownership and inadequate tracking of geographical movements. Building permit data for mobile homes are often of dubious quality.

Separation by type of housing unit in the analysis will not solve all the problems inherent to building permit techniques, of course. Building permit data indicate *intent* to build, not the reality of building. Completion rates and time lags vary with economic conditions, particularly in the case of large multi-unit structures.[1] The conversion of housing units from residential to non-residential use is nearly impossible to monitor and demolitions are generally under-reported. Housing units built for permanent occupancy cannot be differentiated from those built for seasonal occupancy. All these factors add to the imprecision of estimating housing units from building permits.

Even if these problems could be overcome, the problem of estimating current occupancy rates would remain. Direct surveys are the best way to estimate current occupancy rates but the cost of sufficiently extensive surveys is generally prohibitive. Postal vacancy surveys are sometimes used to measure changes in occupancy rates, but their usefulness is limited by a lack of correspondence between postal area and municipal boundaries and by differences in the definition of permanently occupied units. A comparison of postal vacancy rates and census vacancy rates in 1970 for a number of cities in the state of Washington showed census rates to be consistently higher than postal rates; the differences were frequently quite large (Lowe et al., 1977, p. 11).

The problem of estimating current occupancy rates is most commonly dealt with by assuming the rates observed in the most recent census have not changed.

This assumption will generally yield good results when the time interval from the most recent census is short. As the time interval increases, however, occupancy rates become subject to change. In particular, local occupancy rates fluctuate considerably in response to changes in the local and national economy. The reliability of estimates based on the occupancy rates observed in the most recent census thus declines over time.

Starsinic and Zitter found that better estimates of households could be made using a technique based on residential electric customers (Starsinic and Zitter, 1968, p. 477). We have found the same to be true in Florida. Electric customer data are often of better quality than building permit data and are more likely to be available for all places. More important, households can be estimated directly from active residential customers. The intermediate steps of estimating time lags, completion rates, demolitions, conversions and occupancy rates—necessary when estimating households from building permits—are eliminated.

Starsinic and Zitter estimated the net change in households as the net change in the number of active residential electric customers:

$$H_t = H_c + (REC_t - REC_c) \qquad (4)$$

Where:

H_c = occupied housing units in most recent census

REC_t = active residential electric customers at time t

REC_c = active residential electric customers at time of most recent census

We believe households can be more accurately estimated by computing the ratio of households to active residential electric customers at the date of the most recent census and applying this ratio to the number of customers at the estimation date:

$$H_t = (H_c/REC_c) \cdot REC_t \qquad (5)$$

In Florida in 1970 the number of households and the number of active residential electric customers were within 5 percent of each other in only one-third of all local areas. They were more than 20 percent apart in almost one-fourth of the local areas. While these differences are probably larger in Florida than most other states because of Florida's large seasonal population, they are undoubtedly present to some degree in every state. It is highly unlikely that the one-to-one correspondence of increases in households to increases in electric customers implied by equation (4) will hold true in most local areas.

There are a number of reasons for the divergence between households and electric customers. Housing units used seasonally or held for occasional use may have active electric meters but not be occupied by permanent residents. One customer may represent many households if a master meter (one meter serving two or more housing units) is present.[2] Individual meters are often found for pumps, barns, outdoor areas and other types of residential but non-housing use. These meters may show up as separate customers. Bookkeeping practices of individual electric utility companies may differ in the handling of special cases (e.g., a business operating out of a private home may be classified as either a commercial or residential unit). The geographic boundaries used by the electric power company may not correspond exactly to those used by the Bureau of the Census. In addition, there are still a few homes without electricity.

All these factors can lead to differences between the number of households and the number of active residential electric customers. Although a ratio technique takes explicit account of such differences, the validity of this technique rests on the assumption that the ratio of households to residential electric customers remains constant over time. Several factors can induce changes in this ratio. Power companies may provide inaccurate data or change their accounting systems; the use of master meters may increase or decrease; the proportion of housing units used seasonally or held for occasional use may change. In spite of these potentially complicating factors, we have found that ratios generally remain quite stable over time and as will be seen in the following section, special censuses in Florida show the ratio technique to provide better estimates of households than do building permit techniques or the technique which assumes a one-to-one correspondence between changes in electric customers and changes in households.

Seasonality poses perhaps the most difficult problem in estimating permanent households for both building permit and electric customer techniques. There is no way to determine whether a building permit or electric customer represents a permanent household or a seasonal (i.e., non-permanent) housing unit. If the ratio of permanent households to seasonally occupied units is changing over time, both electric customer and building permit techniques will produce inaccurate household estimates. The only way to measure such changes is with sample surveys which are generally too expensive to be feasible. Fortunately most places have very few seasonal units, and for those that do, the seasonal component often remains fairly stable over time. The development of inexpensive, effective indicators of seasonal usage should be a primary goal of future research.

Persons Per Household

Starsinic and Zitter used two techniques for estimating the average number of persons per household (PPH). One technique used the PPH from the most recent census:

$$PPH_t = PPH_c \quad (6)$$

Where:

PPH_t = average number of persons per household at time t

PPH_c = average number of persons per household in most recent census

The other used a linear extrapolation of the trend between the two most recent censuses:

$$PPH_t = PPH_c + x/y\,(PPH_c - PPH_{c-1}) \quad (7)$$

Where:

PPH_{c-1} = average number of persons per household in second most recent census

x = number of years between times c and t

y = number of years between times $c-1$ and c

Neither of these techniques is likely to be very accurate when PPH trends are changing rapidly. Declines in PPH have been much larger during the 1970s than during either of the previous two decades. National PPH declined by 1.2 percent between 1950 and 1960, by 5.7 percent between 1960 and 1970 and by 10.5 percent between 1970 and 1978 (U.S. Bureau of the Census, 1978b, p. 3). When PPH is declining at an increasing rate, techniques based on past censuses or inter-censal trends will tend to overestimate current PPH. This is another cause of the upward bias found in the Starsinic and Zitter population estimates.

We have developed a technique for estimating PPH which incorporates post-censal data as well as data from the most recent census. This technique is based on local PPH in 1970, national trends in PPH since 1970 and estimates of local change in the mix of occupied housing units since 1970. We believe this combination of local and national data has the potential to improve the accuracy of PPH estimates, particularly when a number of years have passed since the previous census or an area is growing very rapidly.

This technique has two components. The first relates national post-censal changes in PPH to the local area. Since estimates of PPH for the United States are available annually from the March Current Population Survey, proportional changes since 1970 can be readily calculated from this data series. Estimates of

PPH for local areas could be made by assuming that the PPH of each local area has declined by the same percentage as the PPH for the nation as a whole, but we believe better estimates can be made by relating this percentage decline to each local area's level of PPH in 1970. There are levels below which PPH will drop no further. As a local area's PPH approaches some lower bound, the percentage decline in PPH must become smaller. Applying the national percentage change in PPH to a local area that already had a low level of PPH in 1970 will therefore tend to underestimate that area's PPH.

We believe it is reasonable to assume that areas with PPH smaller than the nation will exhibit smaller percentage declines in PPH than the nation as a whole while areas with PPH larger than the nation will exhibit larger percentage declines, other things being equal. There is some empirical evidence to support this assumption, as well as the logic described above.[3] This assumption can be quantified in the following relationship:

$$D_a = \left(\frac{PPH_c - L}{PPH_{us} - L}\right) \cdot D_{us} \qquad (8)$$

Where:

D_a = proportional change in PPH for local area since the most recent census

D_{us} = proportional change in PPH for U.S. since most recent census

PPH_c = PPH for local area in most recent census

PPH_{us} = PPH for U.S. in most recent census

L = Lower bound for PPH

The lower bound for PPH (L) has been tentatively set at 1.5. This is the level below which PPH is assumed to drop no further. Although a level of 1.0 is conceptually possible (one person in each occupied housing unit), it is extremely unlikely that such a low level would ever occur in reality. In Florida in 1970 the lowest PPH in any city was 1.76; for counties the lowest was 2.35. A lower bound of 1.5 thus falls between the lowest possible level and the lowest level observed in Florida in 1970. Sensitivity tests indicate that the exact value chosen for L is not crucial to the analysis. Alternate values of 1.25 and 1.75 were found to have very little effect on estimates of PPH.

Equation (8) relates national trends in PPH to estimates of PPH for local areas but does not take account of any factors that might indicate local deviations from national trends. The second component of our technique incorporates an indicator of such deviations by focusing on the postcensal change in the mix of occupied housing units in the local area. The change in housing mix is important because single family, multifamily and mobile home units often have considerably different values of PPH. In 1970 in Florida PPH was 3.11 for single family units, 2.13 for multifamily units and 2.32 for mobile homes. For many cities and counties these differences were even larger. We believe that a change in the mix of occupied housing units in a local area provides an indication of changes in that area's PPH. If, for example, two local areas had identical housing inventories and PPHs in 1970, and one area has added only single family units to its housing inventory while the other has added only multifamily units, we would expect a lower PPH in the latter area than the former, other things being equal.

The effect of changes in local housing mix on PPH can be estimated as follows:

$$PPH_c = \sum_i w_{ic}\, PPH_{ic} \qquad (9)$$

Where:

PPH_c = PPH for all households in most recent census

PPH_{ic} = PPH for i^{th} housing type in most recent census

$w_{ic} = H_{ic}/H_c$

H_{ic} = occupied housing units of type i in most recent census

H_c = total occupied housing units in most recent census

i = single family, multifamily, mobile home units.

$$PPH_t^* = \sum_i w_{it} PPH_{ic} \quad (10)$$

Where:

PPH_t^* = PPH for all households at time t under the assumption that all PPH_i are unchanged since most recent census

$w_{it} = H_{it}/H_t$

H_{it} = occupied housing units of type i at time t

H_t = total occupied housing units at time t.

$$D_m = (PPH_t^* - PPH_c)/PPH_c \quad (11)$$

Where:

D_m = proportional change in PPH due solely to a change in the mix of occupied housing units.

The following identities are implicit to equations (9) and (10):

$$\sum H_{ic} = H_c \quad \text{and} \quad \sum H_{it} = H_t \quad (12)$$

$$\sum w_{ic} = 1 \quad \text{and} \quad \sum w_{it} = 1 \quad (13)$$

Equations (9) and (10) express PPH for all households as a weighted average of the PPH in single family, multifamily and mobile home units. Equation (9) shows the overall PPH at the time of the most recent census and equation (10) shows the overall PPH that would exist at time t if the mix of occupied housing units (w_i) had changed but PPH values by type of housing unit had not. Equation (11) shows the proportional difference between PPH_t^* and PPH_c. This difference is an estimate of the proportional change in overall PPH due solely to the change in the mix of occupied housing units in a local area. We believe this estimate can provide a valuable indication of changes in PPH specific to a given local area. This change in PPH can be either positive or negative, depending on the specific changes in the housing mix of a local area.

While many areas in the United States have undergone substantial shifts in the mix of occupied housing units since 1970, the mix for the United States as a whole has remained quite stable. In 1970, 72.7 percent of U.S. households lived in single unit structures (including mobile homes) and 27.3 percent lived in multi-unit structures (U.S. Bureau of the Census, 1973, p. 45). In 1978, 72.5 percent lived in single unit structures (including mobile homes) and 27.5 percent lived in multi-unit structures (U.S. Bureau of the Census, 1979b, p. 142).[4] Since the proportion of households in single and multi-unit structures changed very little between 1970 and 1978 for the United States as a whole, the 10.5 percent national decline in PPH between 1970 and 1978 cannot be explained by a change in the national mix of occupied housing units. The effects of changes in housing mix for a local area must therefore be considered in addition to—rather than as a part of—the national trend toward smaller PPH.[5]

An estimate of the total proportional change in PPH since 1970 can be expressed as the summation of the proportional change due solely to changes in the mix of occupied housing units (D_m) and the proportional change due to all other factors (D_a):

$$D_t = D_a + D_m \quad (14)$$

Where:

D_t = total proportional change in PPH since most recent census

D_a = proportional change in PPH related to national trends

D_m = proportional change in PPH related to local housing mix

The final step in estimating PPH is to apply the proportional change calculated in equation (14) to the PPH at the time of the most recent census to get an estimate of current PPH for any local area:

$$PPH_t = (1 + D_t)\, PPH_c \quad (15)$$

This technique of estimating PPH is not flawless, of course. It cannot account for differing rates of decline in PPH for different types of housing units. In addition, there are certainly local areas where PPH is moving counter to national trends. We believe, however, that this technique has a great deal of potential for producing accurate estimates of PPH. It is a conceptually simple technique, yet takes account of post-censal trends in national PPH and local housing mix; it incorporates the observation that proportional declines in PPH for areas with high levels of PPH tend to be greater than for areas with low levels of PPH; it uses data available for most local areas. Although the empirical evidence for fully evaluating this technique will not be available until 1980 census results have been tabulated, the evidence now available indicates that it performs quite well in estimating PPH.

The estimate of changes in local housing mix requires current estimates of occupied housing units by type of unit. Such estimates can be obtained from several sources. Direct housing counts provide the best information but are seldom available. Some electric power companies provide data on active residential customers by type of unit, but most are unable to do so. The most widely available source of housing data by type of unit is building permit files. This is the source we have used in preparing our estimates of PPH. Although building permit data are subject to the limitations mentioned earlier, we believe they provide a sufficiently accurate indication of changes in housing mix to be useful in this context. The effect of errors in building permit data on estimates of PPH is generally small because the housing mix component (D_m) of changes in PPH is generally small compared to the component related to national trends (D_a).

We disaggregate housing data into three categories: single family, multifamily and mobile home units. Other disaggregations could be used, such as single family attached units and single family detached units. Disaggregation provides valuable information on what is happening to an area's overall PPH. We chose these three categories because data are readily available from building permit files and PPH differences among the categories are large. The specific categories used for any given place will depend on the availability of data and the magnitude of PPH differences among categories.

Ideally we would estimate PPH separately for single family, multifamily and mobile home units and apply these estimates directly to estimates of households by unit type. Not only would this procedure allow for different rates of PPH decline in each type of housing unit, but it would implicitly account for changes in the mix of housing units. Unfortunately the data needed to estimate PPH by type of unit are not available at the present time. Data on overall PPH are available each year from the Current Population Survey but data on PPH by type of unit are not. Without these data changes in PPH by type of unit cannot be estimated. If such data were available, their use would further improve this technique.

Group Quarters

The number of persons living in group quarters is usually the simplest of the three components of the housing unit method to estimate. (Group quarters include such facilities as college dormitories, military barracks, nursing homes and prisons.) When an area has no large group quarters facilities it is reasonable to assume either that no change in group quarters population has occurred since the previous census or that the ratio of group quarters population to population in housing has not changed. We prefer the latter assumption which implies that group quarters population changes at the same rate as population in housing. Estimates from these two techniques will be very similar unless the local area is growing very rapidly. When large group quarters facilities are present (e.g., barracks, dormitories) the exact number of residents must be obtained directly from administrators of the facilities. This rarely presents a major problem.

EMPIRICAL RESULTS

Data for thoroughly testing these techniques will not be available until after the 1980 census has been tabulated. A number of special censuses have been conducted in Florida during the last several years, however, which provide a basis for limited empirical testing. Although the number is too small to provide conclusive evidence, the results of these censuses provide preliminary evidence that the techniques discussed in this paper improve the accuracy of housing unit population estimates.

Twenty-two special censuses were conducted in Florida between December, 1977 and May, 1979 in cities and unincorporated areas ranging in size from 240 to 123,722. These provide the data for the empirical analysis.[6] It must be emphasized that these censuses were requested and paid for by the local governments of these places. They are not a random sample of places in Florida and may not be a representative sample. Most of these places have grown very rapidly since 1970—eight more than doubled in size. Conjecture that the population of these places would be more difficult to estimate accurately than more slowly growing areas is tempting but difficult to support. It seems unlikely, however, that some factor peculiar to these self-selected places would favor any one of the techniques tested in this paper over the others. For purposes of comparing techniques, the fact that these places were self-selected should present no special problems.

Households

Households were estimated using four different techniques, two based on building permits and two on electric customers. SZ-BP is the Starsinic and Zitter building permit technique defined in equation (2); FLA-BP is our building permit technique defined in equation (3); SZ-REC is the Starsinic and Zitter electric customer technique defined in equation (4); and FLA-REC is our electric customer technique defined in equation (5). Estimates of households produced by these four techniques were compared to the number of households enumerated in the special censuses (adjusted to common dates). The percentage errors calculated from these estimates are summarized in Table 1. The average absolute percentage error provides a measure of the precision of each technique, while the average algebraic percentage error provides a measure of the bias.

Table 1 shows overall precision to be quite similar for the FLA-REC, FLA-BP and SZ-BP techniques. All have overall average absolute errors between 10 and 12 percent, which is much lower than the 22 percent error produced by the SZ-REC technique. For places with greater than 1,000 population in 1970, however, the FLA-REC technique has greater precision than the others. For these places FLA-REC has an average absolute error of 7.5 percent, compared to 11.2 percent for FLA-BP, 10.4 percent for SZ-REC and 11.5 percent for SZ-BP. Differences between FLA-BP and SZ-BP are essentially non-existent because few places in this sample exhibited much change in housing mix between 1970 and 1978. Our technique of separating housing units by type has little effect in such cases.

More dramatic results can be seen in the bottom panel of Table 1. Average algebraic errors are much smaller for FLA-REC than for any of the other three techniques. Overall errors are 2.8 percent for FLA-REC, 5.7 percent for FLA-BP, 18.5 percent for SZ-REC and 8.7 percent for SZ-BP. For places with greater than 1,000 population in 1970 average errors are 3.6, 8.0, 9.1 and 9.6 percent respectively. The positive signs indicate that all four techniques have overestimated occupied housing units.

The results reported in this table support our contentions regarding estimates of households. A technique based on the

Table 1.—Average Percentage Errors for Estimates of Households

Population in 1970	Number of Places	FLA-REC	FLA-BP	SZ-REC	SZ-BP
	Average Absolute Percentage Errors				
<1000	8	20.8	10.9[a]	41.7	7.0[a]
1000-2499	6	6.7	10.5[b]	14.2	11.2[b]
2500+	8	8.1	11.6	7.5	11.7
TOTAL	22	12.4	11.1[c]	21.8	10.2[c]
	Average Algebraic Percentage Errors				
<1000	8	1.4	-0.3[a]	35.0	6.4[a]
1000-2499	6	0.0	6.4[b]	12.5	9.6[b]
2500+	8	6.3	9.0	6.6	9.6
TOTAL	22	2.8	5.7[c]	18.5	8.7[c]

a--Data available for 5 of 8 places

b--Data available for 5 of 6 places

c--Data available for 18 of 22 places

ratio of households to electric customers provides better estimates of households than does a technique based on net change in electric customers or techniques based on building permits. The FLA-REC technique displays greater precision than SZ-REC for all size categories and greater precision than FLA-BP and SZ-BP for places with greater than 1,000 population. More important, it displays a smaller upward bias than any of the other three techniques.

Persons Per Household

Federal special censuses provide information on total population and number of households but generally do not provide breakdowns between population in housing and population in group quarters. Since PPH is defined as population in housing divided by the number of households, exact calculations of PPH are impossible for places in this sample. Population in housing can be estimated, however, by assuming that it represents the same proportion of total population now as it did in 1970. Special census PPH can then be estimated by dividing this estimate of population in housing by the number of households. This provides an excellent estimate of PPH because a majority of the places in the sample had no group quarters population at all in 1970 and no place had more than 1 or 2 percent of total population in group quarters; to our knowledge none of these places has added a substantial group quarters population since 1970. We believe this estimate of PPH provides a very good instrument for testing alternate techniques of estimating PPH.

Three different techniques were used to estimate PPH. SZ-CENSUS simply uses the PPH observed in the 1970 census as an estimator of current PPH, as defined in equation (6). SZ-EXTRAP uses a linear extrapolation of the 1960–1970 changes in PPH, as defined in equation (7). FLA-COM combines local PPH in 1970, national trends since 1970 and changes in local housing mix, as described in equations (8–15). It must be noted that changes in housing mix—the D_m component in equation (14)—cannot be estimated for unincorporated places or cities with less than 2,500 population in 1970. Although building permit and electric customer data are available for these places, 1970 census data on the mix of housing units are not. For these places the FLA-COM estimate of PPH is based solely on PPH in 1970 and national trends since 1970—the D_a component in equation (14).

The percentage errors from these estimates are summarized in Table 2. It is clear from these tables that PPH in 1970 is not a good estimator of current PPH. SZ-CENSUS has an average absolute error of 11.5 percent, almost twice as large as the average errors for the other two techniques. It also has a strong upward bias, as indicated by its average algebraic error of 9.0 percent. It produces an overestimate of PPH in 18 of the 22 places in the sample; in 12 places the estimate is more than 10 percent too high.

A comparison of the other two techniques is made difficult by a lack of data for SZ-EXTRAP. Published data for 1960 are not available for unincorporated balances of counties or cities with less than 1,000 population. Consequently SZ-EXTRAP could be applied to only seven of the 22 places in the sample. On the basis of this limited comparison it appears that FLA-COM produces better estimates of PPH than does SZ-EXTRAP. The level of precision for FLA-COM is slightly better, as indicated by an average absolute error of 5.9 percent compared to 6.8 percent for SZ-EXTRAP. The degree of bias is also considerably smaller. FLA-COM has an average algebraic error of −2.5, compared to 5.1 for SZ-EXTRAP.

Data for both D_a and D_m in equation (14) are available for only seven of the 22 places in the sample. For the other 15 places FLA-COM is based solely on D_a. In order to isolate the effects of changes in housing mix on our estimates of PPH we made one estimate which included housing mix (D_a and D_m) and one which excluded housing mix (D_a only). These results are shown in Table 3. Including the effects of mix change improved the estimate in four of seven places. Furthermore, three places had absolute errors of less than 2 percent when mix change was included, while no place had an error that small when mix change was excluded. The overall results are ambiguous, however. Including mix change lowered the average absolute percentage error somewhat but led to a larger negative bias. Although we believe there are strong reasons for expecting the inclusion of changes in housing mix to improve estimates of PPH, this belief could not be substantiated from this small sample. Conclusive empirical evidence will not be available until tests can be run against 1980 census results.

Population

Estimates of total population were made from several combinations of household and PPH estimation techniques. The errors resulting from these estimates are shown in Table 4. The estimate FLA uses the FLA-REC technique for estimating households and the FLA-COM technique for estimating PPH. The other four population estimates are com-

Table 2.—Average Percentage Errors for Estimates of PPH

Population in 1970	Number of Places	FLA-COM	SZ-CENSUS	SZ-EXTRAP
	Average Absolute Percentage Errors			
<1000	8	6.5	14.0	N/A
1000-2499	6	5.5	8.9	8.3[a]
2500+	8	5.6	11.0	6.2[b]
TOTAL	22	5.9	11.5	6.8[c]
	Average Algebraic Percentage Errors			
<1000	8	0.1	11.5	N/A
1000-2499	6	-2.7	7.9	4.1[a]
2500+	8	-5.0	7.2	5.5[b]
TOTAL	22	-2.5	9.0	5.1[c]

a--Data available for 2 of 6 places

b--Data available for 5 of 8 places

c--Data available for 7 of 22 places

binations using the SZ-BP and SZ-REC techniques to estimate households and the SZ-CENSUS and SZ-EXTRAP techniques to estimate PPH. Although the sample size was too small to demonstrate statistically significant differences among the techniques evaluated in this paper, the FLA technique was found to consistently produce population estimates with smaller errors than the other techniques. The overall average error for the FLA technique was 12.3 percent, compared to 18.0 percent for SZ-BP-CENSUS and 31.3 percent for SZ-REC-CENSUS. The SZ-BP-EXTRAP and SZ-REC-EXTRAP techniques could be applied only to seven places with greater than 1,000 population in 1970. These techniques produced average absolute errors of 12.4 and 12.0 percent respectively, considerably larger than the 5.8 percent error produced by the FLA technique for places of the same size. For places with greater than 2,500 population in 1970, the FLA technique produced an average absolute error of 2.7 percent. This is well within the range of error that has come to be expected for local population estimates and is even below the errors reported by Starsinic and Zitter from their sample of cities with greater than 50,000 population (p. 481).

The results regarding bias are even more dramatic. The FLA technique has very small average algebraic errors for all population size categories. For the entire sample the average algebraic error was

Table 3.—Estimates of PPH and Adjusted Special Census PPH

Place	Adj. Special Census PPH	PPH[a] (D_a)	Percent Error	PPH[b] ($D_a + D_m$)	Percent Error
Ft. Myers	2.53	2.65	4.7	2.53	0.0
Pembroke Pines	2.84	2.76	-2.8	2.62	-7.7
Cape Coral	2.32	2.39	3.0	2.34	0.9
Dania	2.24	2.45	9.2	2.28	1.8
Port Orange	2.40	2.30	-4.4	2.20	-8.3
Edgewater	2.42	2.33	-3.6	2.30	-4.8
Cooper City	3.19	2.49	-21.9	2.54	-20.1
Average Absolute Percent Error			7.1		6.2
Average Algebraic Percent Error			-2.3		-5.5

a--Estimate of PPH excluding effects of changes in housing mix.

b--Estimate of PPH including effects of changes in housing mix.

−0.1 percent for FLA, compared to 16.8, 11.7, 27.0 and 9.4 percent for the other four techniques. This very low average algebraic error indicates an almost complete lack of bias for the FLA technique in the present sample. This contrasts sharply with the strong upward bias exhibited by the other techniques.

There is a strong negative relationship between size of place and the precision of estimates. Average absolute errors for the FLA technique were 23.7 percent for places with less than 1,000 population in 1970, 9.9 percent for places 1,000–2,499 and 2.7 percent for places 2,500 or greater. Such a result is not uncommon. Variations around national trends are greater for small places than large places, and the effects of absolute data errors are magnified in small places. Growth rates are often high as well, adding instability to the population base. Three of the eight places in the sample with less than 1,000 population in 1970 tripled in population between 1970 and the date of the special census. The average absolute error for these places was 37.6 percent; for the other five places it was 15.6 percent.

DISCUSSION

The techniques discussed in this paper were tested using special census data collected in 22 places in Florida. Although this sample is considerably too small to provide conclusive comparisons, it does provide preliminary evidence that estimates produced by the techniques we have developed are more precise (i.e., smaller average absolute percentage errors) and less biased (i.e., smaller average algebraic percentage errors) than estimates produced by the other techniques.

The apparent success of these techniques, of course, does not imply that all problems have been solved. There are many areas in which further improve-

Table 4.—Average Percentage Errors for Estimates of Total Population

Population in 1970	Number of Places	FLA	SZ-BP CENSUS	SZ-BP EXTRAP	SZ-REC CENSUS	SZ-REC EXTRAP
Average Absolute Percentage Errors						
<1000	8	23.7	20.1[a]	N/A	50.7	N/A
1000-2499	6	9.9	15.9[b]	5.5[d]	24.6	18.7[d]
2500+	8	2.7	18.0	15.2[e]	16.8	9.3[e]
TOTAL	22	12.3	18.0[c]	12.4[f]	31.3	12.0[f]
Average Algebraic Percentage Errors						
<1000	8	2.0	15.6[a]	N/A	46.9	N/A
1000-2499	6	-2.5	15.9[b]	5.5[d]	21.9	9.6[d]
2500+	8	-0.3	16.6	14.2[e]	11.0	9.3[e]
TOTAL	22	-0.1	16.8[c]	11.7[f]	27.0	9.4[f]

a--Data available for 5 of 8 places

b--Data available for 5 of 6 places

c--Data available for 18 of 22 places

d--Data available for 2 of 6 places

e--Data available for 5 of 8 places

f--Data available for 7 of 22 places

ment is possible. The quality and consistency of building permit and electric customer data can be improved. New sources of household data can be explored. New measures of occupancy rates can be found. Indicators of changes in seasonality can be constructed. New techniques for detecting local deviations from national PPH trends can be developed. We believe the techniques discussed in this paper represent an improvement in the state of the art of housing unit population estimates, but the limit to further improvements certainly has not yet been reached.

The techniques we have developed are not based on data unique to the state of Florida. They can be applied in any region of the United States and can be adapted to utilize different sources of data that may exist in various local areas. Moreover, they can be applied at relatively little cost since they are based on widely available data and require no extensive surveys or special censuses. The effort needed to apply these techniques is thus within the reach of most agencies making local population estimates, even those with small staffs and limited budgets.

These techniques do, however, require accurate base data. Consistent and comprehensive building permit and/or electric customer data are essential. Accurate data corresponding to the date of the decennial census are particularly important because they form the benchmark relationships upon which all future estimates are based. As with any other estimation technique, if the base data are of poor quality the population estimates will be of poor quality as well.

We believe accurate data can be collected if sufficient effort is expended. Our experience in collecting data in Florida has been quite positive. We have been able to gather electric customer data for every city and county in the state. All 55 electric power companies in the state provide us their data; only one company charges for this service. Companies can be motivated to provide electric customer data not only because doing so promotes a favorable image with the public, but also because accurate population estimates are of great importance to them as planning tools. The data provided by large companies are typically of very good quality; data provided by small companies (particularly cooperatives) are much more likely to be inconsistent or erroneous. Small companies are also less likely to be able to accurately disaggregate customer data by geographic boundaries. Since large companies serve the main population centers of the state, however, the overall quality of electric customer data is very high. Building permit data are also generally of good quality, except for mobile homes. While not available everywhere, we have found them to be available for the large majority of places in the state. Places without building permit data are almost always quite small; their effect on the total estimation program is therefore minimal.

One frustration we have experienced in applying the techniques described in this paper is the difficulty of obtaining census data on PPH by type of housing unit. These figures are not available in printed form, but must be computed from data found only on census tapes. The publication of these figures in decennial and special census reports would be a valuable tool for the many agencies applying housing unit techniques. Agencies with limited budgets would be particularly benefitted.

A similar problem relates to the Current Population Survey. The estimate of national PPH from the March survey is essential to our technique of estimating local PPH. National estimates of PPH by type of housing unit would also be most useful. These estimates would permit the calculation of separate PPH declines for each type of housing unit. Although the necessary data have been collected since 1975, they are available neither in printed form nor in the computer tape file released to the public. Making these data easily available would be a valuable service to agencies using housing unit techniques.

Research on methods of estimating the population of local areas has focused largely on component and regression methods (e.g., Goldberg et al., 1964; Zitter and Shryock, 1964; Rosenberg, 1968; Pursell, 1970; Namboodiri, 1972; O'Hare, 1976). Only a few studies have focused on the housing unit method (e.g., Voss and Krebs, 1979; Findley and Reinhardt, 1980). This apparent imbalance in research effort is puzzling because the housing unit method is the most widely used method of estimating local population and is one of very few methods capable of producing subcounty population estimates. Starsinic and Zitter concluded that their application of the housing unit method performed surprisingly well in producing local population estimates. The present study also found housing unit techniques to produce fairly accurate estimates. Research directed toward further improvements in the application of housing unit techniques promises a large payoff in improved population estimates, particularly for subcounty areas.

CONCLUSIONS

The housing unit method has long been treated as the ugly step-sibling in the family of population estimation methods. The general sentiment among demographers seems to be that the housing unit method is better than nothing, but is not the equal of component, regression or administrative records methods of population estimation. It is time that this sentiment be challenged. The housing unit method is capable of producing extremely accurate population estimates. It is particularly valuable for estimating subcounty populations which cannot be estimated with most other methods because of data limitations. With further research into new sources of data and new techniques of application, and with thorough testing against census results, the housing unit method can become at least the equal to any other method for estimating local population.

NOTES

[1] Surveys can help solve this problem. In Florida we periodically contact the purchasers of permits for large multi-unit structures and collect information on the current status of construction. This information allows us to adjust our data for unfinished structures. Surveys also provide valuable information on average time lags.

[2] Starsinic and Zitter note that master metered units are becoming more common over time (p. 477). While this may have been true in the 1960s, the opposite has been true in Florida in the 1970s. Few master metered units have been built and many units formerly on master meters have been converted to individual meters. With the recent major increases in energy prices, it is likely that master metered units will become less and less common.

[3] A linear correlation coefficient for PPH in 1960 versus the percentage decline in PPH between 1960 and 1970 was computed for Florida's 67 counties. This correlation was positive (0.34) and significant at .005. While this is a fairly weak test because no other variables were controlled for, it does provide some empirical justification for our assumption that places with high PPH will exhibit larger percentage declines in PPH than places with low PPH.

[4] Current data on the number of households occupying mobile homes are not available. The effect of changes in mobile homes on national housing mix cannot be great, however, because mobile homes account for a very small percentage of permanently occupied housing units (3.3 percent in 1970).

[5] If the national mix of housing units changes over time, this change must be accounted for in calculating the effects of local mix change on local PPH. Specifically, national decline in PPH must be separated into two parts, one related to changes in national housing mix and the other related to all other factors. Only the second part can be used as D_{us} in equation (8). The estimate of D_{us} used in equation (8) should be independent of any change in the national mix of housing units.

[6] Population estimates in Florida are for July 1 of each year. All special census results are adjusted backward to the previous July 1 to make them comparable to the estimates. These adjustments are made by interpolating trends in number of households and PPH. All tables in this paper are based on the adjusted special census data. Complete special census results for each city and unincorporated area are available from the authors upon request.

[7] Standard statistical tests for differences in average algebraic percentage errors were run for each technique's estimates of households, PPH and population, for each size-of-place category. The hypothesis that the average algebraic percentage error was equal to zero could not be rejected at the .10 level for any of the estimation techniques. In addition, no significant differences (at .10) were found between either the means or the variances of the algebraic percentage errors observed for these techniques. We believe this lack of significance was caused by the small sample size.

ACKNOWLEDGMENTS

The authors are indebted to the Bureau of Economic and Business Research of the University of Florida for providing financial support for this study. They also thank two anonymous referees for their helpful suggestions and Melinda Craine and Donna Carman for valuable research and clerical assistance. An earlier version of this paper was presented at the annual meeting of the Southern Regional Demographic Group, Myrtle Beach, SC, October 17–19, 1979.

REFERENCES

Findley, Sally and Hazel Reinhardt. 1980. Nonlinear Estimation of Household Size: The Minnesota Housing Unit Method. Paper presented at the annual meetings of the Population Association of America, April 10–12.

Goldberg, D., V. R. Rao and N. K. Namboodiri. 1964. A Test of the Accuracy of Ratio Correlation

Population Estimates. Land Economics 40:100–102.
Lowe, Theresa, Donald Pittenger and John Walker. 1977. Making the Housing Unit Population Estimation Method Work: A Progress Report. Paper presented at the annual meetings of the Population Association of America, St. Louis, Mo.
Namboodiri, N. K. 1972. On the Ratio-Correlation and Related Methods of Subnational Population Estimation. Demography 9:443–453.
O'Hare, W. P. 1976. Report on a Multiple Regression Method for Making Population Estimates. Demography 13:369–379.
Pursell, D. E. 1970. Improving Population Estimates with the Use of Dummy Variables. Demography 7:87–91.
Rosenberg, Harry. 1968. Improving Current Population Estimates Through Stratification. Land Economics 44:331–338.
Siskind, David. 1980. Housing Construction Data Vital. Data User News. Vol. 15, no. 2. February. Washington, D.C.: U.S. Bureau of the Census.
Starsinic, D. E. and M. Zitter. 1968. Accuracy of the Housing Unit Method in Preparing Population Estimates for Cities. Demography 5:475–484.
U.S. Bureau of the Census. 1973. 1970 Census of Housing. Subject Reports HC(7)-1, February.
———. 1978a. Current Population Reports, Series P-25, no. 723, pp. 1–80.
———. 1978b. Current Population Reports, Series P-20, no. 327, pp. 1–7.
———. 1979a. Current Housing Reports, Series H-111, 79-Q2, pp. 1–15.
———. 1979b. Current Population Reports, Series P-20, no. 340, pp. 1–160.
Voss, P. R. and H. C. Krebs. 1979. The Use of Federal Revenue Data for Improving Current Estimates of Average Household Size for Minor Civil Divisions: An Evaluation. Technical Series 70-5. Madison: Applied Population Laboratory, Department of Rural Sociology, University of Wisconsin.
Zitter, M. and H. S. Shryock, Jr. 1964. Accuracy of Methods of Preparing Postcensal Population Estimates for State and Local Areas. Demography 1:227–241.