DEMOGRAPHY© Volume 20, Number 3 August 1983

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

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

Bart B. Lewis
Atlanta Regional Commission, 100 Edgewood Avenue, NE, Suite 1801, Atlanta, Georgia 30335

Abstract—The housing unit method of population estimation is often characterized as being imprecise and having an upward bias. In an earlier paper we argued that the method itself cannot be properly characterized by a particular level of precision or direction of bias. Only specific techniques of applying the method can have such characteristics. In that paper we presented several new techniques for estimating the number of households and average number of persons per household (PPH). However, the testing of these new techniques was limited by the lack of census results against which the estimates could be compared. Complete census data on population, households, and PPH are now available and can be used to test alternate estimation techniques. In this paper we replicate the tests reported in our earlier paper using 1980 census data for Florida's 67 counties. These tests provide further evidence that the new techniques produce more precise, less biased estimates than previously used techniques.

The housing unit method of local population estimation is conceptually clear and theoretically sound. The population of any geographic area is equal to the number of permanently 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, military barracks, penal institutions). If these three components were known exactly, the exact total population also would be known. Unfortunately, each of these components is generally unknown and must be estimated using imperfect techniques and data sources. Consequently errors can enter the estimation procedure.

Many different techniques and data sources have been used to estimate the three components of the housing unit method. Perhaps the most commonly used techniques are those described and evaluated by Starsinic and Zitter (1968). Over the years, these techniques have become virtually synonymous with the housing unit method. In an earlier paper (Smith and Lewis, 1980) we pointed out that these techniques do not intrinsically define the housing unit method, but rather represent specific ways of applying the method. In that paper we described several techniques that we believe will produce more accurate estimates than the techniques evaluated by Starsinic and Zitter. Using special census results from 22 places in Florida, we compared the estimation errors from our techniques with those from the Starsinic and Zitter techniques. We found that our

techniques produced population estimates that were more precise and less biased than those produced by the other techniques.

The validity of our conclusions, however, was weakened by the small number of observations and the self-selective nature of the sample (special censuses are conducted only in places that request and pay for them). In the present paper we replicate the tests described in our earlier paper using data from the 1980 census for all 67 counties in Florida. These data provide a much larger sample size and are free from any bias caused by self-selection. The results of this analysis confirm our original results and strengthen the conclusions drawn in our earlier paper.

EMPIRICAL RESULTS

We compare four techniques for estimating households. Two are based on building permits, one using total permits and demolitions (SZ-BP) and the other differentiating permits and demolitions by type of housing unit (FLA-BP). Two are based on residential electric customers, one using absolute changes in customers (SZ-REC) and the other using a ratio of households to customers (FLA-REC). We also compare three techniques for estimating PPH. One is the PPH existing at the most recent census (SZ-CENSUS); one is a linear extrapolation of the trend between the two most recent censuses (SZ-EXTRAP); and one is derived from the local PPH at the time of the most recent census, the national change in PPH since that census, and the local change in the mix of housing units since that census (FLA-COM). Following the notation of our previous paper, SZ refers to techniques tested by Starsinic and Zitter and FLA refers to the techniques we have developed. A complete description of the data and the assumptions underlying these techniques can be found in Smith and Lewis (1980, pp. 323–339).

Data for evaluating these techniques have been collected through the 1980 census. Enumerations of total population, households, average number of persons per household, and population in group quarters have been published for states and counties (U.S. Bureau of the Census, 1981). In this section we compare the performance of the techniques for estimating households, PPH, and total population, focusing on precision, stability, and bias. Estimates for the group quarters population are not reported, as they represent a very small proportion of total population in most areas and have very little effect on the accuracy of total population estimates.

Households

Table 1 summarizes the errors for estimates of households.[1] Mean absolute percentage error is the mean when the sign of the error is ignored. It provides a measure of the precision of each technique. Mean algebraic percentage error is the mean when the sign of the error is included. It provides a measure of the bias. The standard deviation provides a measure of the dispersion of errors around the mean, or the stability of the estimating technique.

Of the two building permit techniques, FLA-BP produces slightly better estimates than SZ-BP. FLA-BP has smaller mean absolute percentage errors in all four size-of-place categories, but the differences are very small. FLA-BP also has smaller mean algebraic percentage errors than SZ-BP, but again the differences are very small. One cannot conclude from this evidence that FLA-BP produces significantly better estimates of households than SZ-BP.

Of the two electric customer techniques, FLA-REC produces far better estimates than SZ-REC. FLA-REC has much smaller mean absolute percentage errors in all four size-of-place categories, and smaller standard deviations as well. FLA-REC also has much smaller mean algebraic percentage errors, with smaller standard deviations in three of the four

Table 1.—Mean Percentage Errors for Estimates of Households

Population in 1970	Number of Places	SZ-BP	FLA-BP	SZ-REC	FLA-REC
			Mean Absolute Percentage Errors		
<15,000	25	15.03	14.20	12.94	7.81
		(11.04)	(11.53)	(10.45)	(6.00)
15,000-49,999	18	17.41	16.86	9.26	7.17
		(10.22)	(9.57)	(7.76)	(3.77)
50,000-99,999	9	10.97	10.85	11.53	8.17
		(10.15)	(9.71)	(6.53)	(5.87)
100,000+	15	5.33	4.71	6.89	5.82
		(4.94)	(3.89)	(6.82)	(5.53)
Total	67	12.95	12.34	10.41	7.24
		(10.47)	(10.34)	(8.72)	(5.30)
			Mean Algebraic Percentage Errors		
<15,000	25	-12.22	-10.57	11.98	4.21
		(14.20)	(15.06)	(11.58)	(9.00)
15,000-49,999	18	- 9.91	- 9.93	6.41	-0.98
		(17.93)	(16.97)	(10.36)	(8.22)
50,000-99,999	9	- 5.51	- 5.76	11.08	5.95
		(14.30)	(13.76)	(7.34)	(8.35)
100,000+	15	4.36	3.09	5.72	3.55
		(5.88)	(5.36)	(7.90)	(7.31)
Total	67	- 6.98	- 6.69	8.96	2.90
		(15.18)	(14.70)	(10.22)	(8.53)

NOTES: SZ-BP—Starsinic and Zitter building permit technique
FLA-BP—Florida (authors') building permit technique
SZ-REC—Starsinic and Zitter residential electric customers technique
FLA-REC—Florida (authors') residential electric customers technique

Standard deviations are given in parentheses.

size-of-place categories. Looking at the total sample, the mean absolute errors are 7.2 percent for FLA-REC and 10.4 percent for SZ-REC, with standard deviations of 5.3 and 8.7 respectively. The mean algebraic errors are 2.9 percent for FLA-REC and 9.0 percent for SZ-REC, with standard deviations of 8.5 and 10.2 respectively. It is clear from these results that FLA-REC produces estimates of households that are more precise, more stable, and less biased than those produced by SZ-REC.

Table 1 appears to show that the FLA electric customer technique produces better estimates of households than the FLA building permit technique. Mean absolute and algebraic percentage errors and standard deviations are much smaller for FLA-REC than for FLA-BP. The apparent superiority of the FLA-REC technique may be spurious, however. A number of counties in Florida do not have complete building permit data. Some local areas do not issue building permits; others have begun to issue them only recently. The largest absolute and algebraic errors are in counties with fewer than 50,000 people. These are the counties with the largest deficiencies in

building permit data. For counties with 100,000 or more people, building permit data are generally nearly complete. In these counties the estimates of households produced by FLA-BP are very similar to those produced by FLA-REC. The large errors and downward bias for FLA-BP shown in Table 1 therefore may be due to incomplete data rather than to the nature of the technique itself.

It is interesting that electric customer estimates of households have an upward bias while building permit estimates of households have a downward bias. The downward bias of the building permit estimates is most likely caused by the lack of complete data in many counties. The upward bias in electric customer estimates is most likely caused by increased seasonality and tourism. If housing units occupied by nonpermanent residents (e.g., tourists, seasonal residents) are increasing as a proportion of total units, the two electric customer techniques described in this paper will tend to overestimate households. This has been occurring in Florida in recent years. The proportion of housing units occupied by permanent residents declined from .904 in 1970 to .855 in 1980. In counties and states where tourism and seasonality are holding relatively constant, the electric customer techniques will likely produce smaller algebraic and absolute errors than those shown here.[2]

Persons Per Household

Errors in the estimates of PPH are shown in Table 2. Using the PPH from the previous census (SZ-CENSUS) clearly produces the worst estimates. SZ-CENSUS has the largest mean absolute percentage error in every size-of-place category and has a strong upward bias, as indicated by the large positive mean algebraic percentage errors. The extrapolation technique (SZ-EXTRAP) produces better estimates of PPH than SZ-CENSUS, but does not perform as well as the FLA-COM technique. The FLA-COM technique has smaller mean absolute errors and standard deviations than SZ-EXTRAP for all four size-of-place categories. The mean absolute percentage error for the entire sample is 3.7 for FLA-COM and 6.5 for SZ-EXTRAP. These errors show the FLA-COM technique to be much more precise than SZ-EXTRAP. Variation around the mean is also much smaller, as shown by the standard deviations of 2.5 for FLA-COM and 9.2 for SZ-EXTRAP. In terms of bias, however, the techniques show smaller differences. The FLA-COM technique has a slight tendency to underestimate PPH, and the SZ-EXTRAP technique has a slight tendency to overestimate.

Estimates of households and PPH are the two primary components of the housing unit method. Which component can be estimated more accurately? A comparison of the errors for FLA-REC in Table 1 and FLA-COM in Table 2 shows that estimates of PPH are more precise and more stable than estimates of households. Average absolute percentage errors and standard deviations are smaller for FLA-COM than for FLA-REC in every size-of-place category. For the total sample, the average error and standard deviation are only half as large for FLA-COM as for FLA-REC. These larger errors for estimates of households than for estimates of PPH are most likely due to the much higher rate of growth for households during the decade. Between 1970 and 1980 the number of households in Florida increased by 64 percent. In 13 counties the number of households more than doubled. On the other hand, PPH declined by only 12 percent in Florida between 1970 and 1980. In no county was the change in PPH greater than 20 percent. The potential for error was therefore much greater for estimates of households than for estimates of PPH.

We would speculate that, in general, it is likely that in places that are growing or declining rapidly, estimates of households will add more to overall population estimation error than will estimates of

Table 2.—Mean Percentage Errors for Estimates of Persons Per Household

Population in 1970	Number of Places	SZ-CENSUS	SZ-EXTRAP	FLA-COM
		Mean Absolute Percentage Errors		
<15,000	25	12.36 (3.91)	5.61 (4.81)	4.44 (2.48)
15,000-49,999	18	12.29 (3.38)	8.31 (16.24)	3.16 (2.50)
50,000-99,999	9	13.71 (5.57)	6.67 (4.29)	4.14 (3.02)
100,000+	15	14.21 (4.47)	5.75 (5.63)	2.85 (1.70)
Total	67	12.93 (4.15)	6.51 (9.20)	3.70 (2.46)
		Mean Algebraic Percentage Errors		
<15,000	25	12.36 (3.91)	1.05 (7.01)	-3.82 (3.39)
15,000-49,999	18	12.29 (3.38)	-2.39 (18.19)	-2.86 (2.86)
50,000-99,999	9	13.71 (5.57)	2.73 (7.75)	-1.21 (5.17)
100,000+	15	14.21 (4.47)	5.07 (6.32)	-2.05 (2.67)
Total	67	12.93 (4.15)	1.25 (11.23)	-2.81 (3.45)

NOTES: SZ-CENSUS—Persons per household from the most recent census.
SZ-EXTRAP—Extrapolation of trend between the two most recent censuses.
FLA-COM—Derivation from local PPH at the time of the most recent census, national change in PPH since that census, and changes in the mix of local housing units.

Standard deviations are given in parentheses.

PPH. In places where the number of households is not changing rapidly, however, this may not be the case. Further study on the source of error is needed.

Population

Estimates of total population were made from several combinations of household and PPH estimation techniques. The errors for these estimates are shown in Table 3. The estimate FLA uses the FLA-REC technique to estimate households and the FLA-COM technique to estimate PPH. The other four population estimates are combinations using the SZ-BP and SZ-REC techniques to estimate households and the SZ-CENSUS and SZ-EXTRAP techniques to estimate PPH. All five techniques use the same estimate for the number of persons living in group quarters.

The FLA technique consistently produces better estimates of total population than do the other four techniques. The mean absolute percentage error is much smaller for FLA than for the other techniques, 7.2 compared to 20.6, 13.8, 14.1, and 14.3. The standard deviation is smallest, 5.5 compared to 12.2, 12.4, 8.0,

Table 3.—Mean Percentage Errors for Estimates of Population

Population in 1970	Number of Places	FLA	SZ-REC		SZ-BP	
			CENSUS	EXTRAP	CENSUS	EXTRAP
			Mean Absolute Percentage Errors			
<15,000	25	6.92	23.74	16.20	14.01	16.57
		(5.46)	(14.09)	(12.06)	(8.25)	(13.62)
15,000-49,999	18	7.48	17.41	13.95	14.53	19.89
		(4.78)	(12.29)	(17.32)	(9.09)	(18.08)
50,000-99,999	9	10.13	23.81	12.86	12.89	9.58
		(7.95)	(8.33)	(7.47)	(7.54)	(12.11)
100,000+	15	5.59	17.44	10.35	14.53	6.55
		(4.44)	(9.36)	(7.56)	(6.97)	(4.84)
Total	67	7.20	20.64	13.84	14.12	14.28
		(5.51)	(12.18)	(12.41)	(7.97)	(14.20)
			Mean Algebraic Percentage Errors			
<15,000	25	0.36	21.14	9.16	- 3.91	-13.22
		(8.92)	(17.89)	(18.20)	(16.02)	(17.01)
15,000-49,999	18	- 3.74	16.99	2.15	- 1.30	-13.26
		(8.20)	(12.90)	(22.39)	(17.45)	(23.66)
50,000-99,999	9	5.44	23.81	11.79	4.96	- 4.96
		(12.06)	(8.33)	(9.24)	(14.70)	(14.91)
100,000+	15	1.86	17.44	8.09	14.53	5.28
		(7.04)	(9.36)	(10.09)	(6.97)	(6.28)
Total	67	0.27	19.55	7.39	2.10	- 7.98
		(9.11)	(13.88)	(17.12)	(16.17)	(18.55)

NOTE: Standard deviations are given in parentheses.

and 14.2. The mean algebraic percentage error is also smallest, .3 compared to 19.6, 7.4, 2.1 and −8.0. The FLA estimates of total population are clearly more precise, more stable, and less biased than the estimates produced by the other four techniques.

SUMMARY AND CONCLUSIONS

The results shown in this paper confirm the findings of our earlier paper. The techniques we have developed for estimating households and PPH perform considerably better than the techniques evaluated by Starsinic and Zitter. The FLA-REC technique for estimating households is more precise, more stable, and less biased than the other techniques. The FLA-COM technique for estimating PPH is more precise and stable than SZ-CENSUS and SZ-EXTRAP and much less biased than SZ-CENSUS. For estimates of total population, the superiority of the FLA techniques is even more evident. When compared to the other techniques, the FLA estimates of total population have greater precision, exhibit less variation around the mean, and have virtually no bias.

The housing unit method is the most commonly used method for making local population estimates in the United States. Yet the method is widely characterized as being inaccurate and having an upward bias. We believe any such characterization is improper. The housing

unit formula itself is a mathematical identity and has no inherent characteristics relating to precision or bias. Only specific techniques of applying the method can have such characteristics. We have demonstrated in this paper that the new techniques we have developed are more precise and less biased than other commonly used techniques. Future research will undoubtedly produce further refinements or completely new techniques that will be even better than those discussed in this paper. The evolution of the housing unit method is by no means complete. We are confident that continued research and development will make the housing unit method an increasingly accurate and useful method of local population estimation.

NOTES

[1] Differences between estimates and census counts can be caused by either estimation error or enumeration error. In this paper we will refer to these differences as errors of the estimates.

[2] Seasonality poses a serious problem for the housing unit method. Sample surveys can alleviate the problem, but are expensive. Research directed toward developing symptomatic indicators of seasonality may lead to very useful refinements to the housing unit method.

ACKNOWLEDGMENTS

The authors are indebted to the Bureau of Economic and Business Research at the University of Florida for providing financial support for this study. An earlier version of this paper was presented at the annual meeting of the Southern Regional Demographic Group, Little Rock, Arkansas, October 14–16, 1981.

REFERENCES

Smith, Stanley K., and Bart B. Lewis. 1980. Some New Techniques for Applying the Housing Unit Method of Local Population Estimation. Demography 17:323–339.

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. 1981. 1980 Census of Population: Supplementary Report. PC80-S1-2, May. Population and Households by States and Counties: 1980. Washington, D.C.: U. S. Government Printing Office.