EFFECTIVENESS OF OUR TOOLS FOR ESTIMATING POPULATION CHANGE IN SMALL AREAS

By: Carl M. Frisen
California Department of Finance

The widespread interest in demographic data relating to small areas is reflected in the number of agencies listed in "Local Population Estimates Prepared by State and City Agencies: 1957-58", published by the United States Bureau of the Census in its Current Population Reports, Series P-25, No. 178, in June 1958. A summary is provided of the estimating work done by 62 state agencies and 33 city agencies. Also included is a useful discussion of the problems and limitations involved in the preparation of small area estimates. From the information included, it is apparent that the users of population data, involving both governmental agencies and private organizations, have been able to stimulate a great deal of activity aimed at providing population information on a current basis. As a result, county population estimates are now generally available and in many states the populations of cities and towns are estimated at regular intervals.

From the standpoint of the producer of demographic materials, problems of method - such as the choice of the most effective tool in a given instance - depend upon the questions posed by the user of the data. For the most part, it is likely that these questions will call for one or more of the following types of information:

1) estimates of total or civilian population, for one or more small areas;
2) measures of the components of change, particularly the volume of net civilian migration;
3) analysis of the composition of population with the concern likely to involve age, sex and color or race;
4) estimates of the number of households, and factors relating to the formation of households and families since the last federal census.

A second consideration relates to the nature of the small areas for which these types of data may be needed. If the State is considered the largest "small area" in this discussion, administrative needs may range from state, counties and cities, through less widely recognized but formally delineated units such as hospital districts, school districts and townships, to such amorphous areas as unincorporated places, marketing areas and the like.

From the standpoint of the producer, then, the administrator's request for demographic information poses problems in terms of the nature of the data that are desired and in terms of the specific areas involved. This means that the demographer must first determine whether symptomatic data are available that can measure the factor with which he is concerned, be it total population, net migration, age distribution or whatever, and secondly he must ascertain whether the symptomatic data are available for the geographic areas designated. Once the availability and accuracy of the indicators are determined, there remains the problem of their applicability in terms of technique.

At this point it is deemed appropriate to suggest two classificatory schemes that may contribute perspective in analyzing the current picture. The first of these is applicable to the symptomatic data used in the estimating process, while the second is concerned with how the estimating technique relates to components of population change.

The indicators selected by various estimators may be classified on the basis of the degree to which they relate to the population that they seek to measure. One class of variables may be termed indirect, since measurements are expressed in units other than persons. Data on utility connections, rural route box holders and occupied or total dwelling units are of this type, for they require a factor for the average number of persons per unit before an estimate of population can be made. A second class of variables includes those that are direct, in that they are measured directly in terms of numbers of people. A distinction may be made in this class between indicators that are partial and total. Partial-direct variables are limited to certain classes or elements of the population and the relationship of the subgroup to the total must be established prior to their use in estimating. Among the variables of this type are school enrollment data, selective service registration, birth statistics and motor registrations. Total-direct variables are complete in their coverage, relating to all segments of the population and include morbidity and mortality records. Illness, accidents and deaths occur at all ages, in both sexes and among all races.

An alternative classification may be made in terms of how the estimating techniques measure population change. Some methods, such as the Census Bureau's methods I and II, estimate only net civilian migration, and other components of change must be obtained separately and combined with the migration estimate to indicate total population change. Other methods provide a measure of total population change only, and net migration may be derived only as residual by subtracting other components. Bogue's vital rates method is of this type, as is the dwelling unit method currently being used in California's city population estimating program. A third category, which may be termed composite, includes estimates based on combinations of methods some of which measure net migration only while others estimate total change. One such combination has been used by Bogue and Duncan in estimating the population of cities in Illinois by age, sex and color. Here, school enrollment is used for the younger ages, fertility ratios for the middle ages, and specific death rates for the older ages.

As modified by the Census Bureau for purposes of a test against the 1950 census, reported on in a paper by Shryock, Siegel and Greenberg read at the Population Association meeting in May 1957, the Census Bureau's method II was used to estimate preschool and school-age migration, while the 18-44 year old change and the 45 year and over change were estimated as total population change.

For the moment, let us direct our attention to the questions facing the demographer who is assigned a specific problem for solution. Many of the analyses of estimating methods fail to give sufficient emphasis to population estimates in terms of their role in the administrative process. The effect of this role may be illustrated in the case of various requests for city population estimates in California. From time to time State agencies ask for current estimates for one or more incorporated places in the State, perhaps for use in making a general evaluation of changes in agency workloads. Ordinarily the Department of Finance will depend upon extrapolations from special censuses or locally published estimates as sufficiently accurate under these circumstances. The problem is of a different order when the question is posed: can satisfactory population estimates for cities be prepared that will serve as the basis for allocating State-collected funds on a per capita basis? This question was raised by the League of California Cities in exploring alternatives to the use of special censuses in officially determining population change. It is readily apparent that in the requests cited we are concerned with two different definitions of "satisfactory" as applied to population estimates for cities, and the demographer must be aware of these differences and what they mean in terms of data gathering and method. In the case of the study for the League of California Cities, an effort was made to furnish a criterion for determining whether a given estimate was satisfactory by approximating the financial effect of consistently underestimating a city's population as compared with the funds received under procedures then currently in use by the city. That is, an estimating technique might be deemed unsatisfactory if a likelihood existed that some cities would receive smaller amounts of allocated funds over a period of time than would be true under the procedure previously in effect.

Before any test of accuracy or acceptance can be applied, however, the producer of population estimates must determine what tools are at hand for the task assigned by the administrator. The first question to arise is: what symptomatic data are available for the specific areas involved? When cities are the object of study, a number of possible indicators may be explored. Records of elementary school enrollment are sometimes available, though in California few cities have school district boundaries that coincide with or even approximate city boundaries. In many instances current information can be obtained from school records relatively easily, but it is not possible to derive data for an earlier census date that must serve as the benchmark. Statistics on births and deaths are tabulated annually for most of the cities in the State, with detailed data, such as deaths by age, sex and race and births by age of mother, compiled for cities of 50,000 and over. Again, local building permit records on residential construction are maintained by most California cities, though there are differences in the completeness of coverage and the time span covered by the records. Finally, information on residential meter connections can be obtained from utility records, though there may be some problems of interpretation of data maintained by some municipal utilities. This brief summary of the major types of indicators found in seeking to establish a broad test of city population estimating techniques shows that it is possible to find some instances of all three classes of symptomatic data, and in turn to utilize them through all three classes of method.

Looking first at partial-direct variables, the present discussion is limited to the use of reported school-age population rather than school enrollment. Special censuses report population by age for cities of 50,000 and over and for a limited number of smaller cities that contract for the detailed information. Since the Census Bureau's methods I and II use the school enrollment to estimate related population age groups, it is feasible to substitute the appropriate population directly in the methods, thus testing their ability to estimate net civilian migration for all ages. Through this procedure, it is possible to avoid problems of availability and accuracy of school enrollment data.

Questions of this order should be explored thoroughly, however, if only to assure that the results will receive proper interpretation. For example, the City of Los Angeles has been included in the published results of a number of tests of method II as applied to large cities in the United States. A proper check in advance would have shown that elementary school enrollment data for the Los Angeles school district cover an area approximately 20 percent larger in population than the total within the corporate limits of Los Angeles. In actuality, it is not possible at the present time to employ only a school enrollment-based method to estimate the population of the City of Los Angeles, and it is preferable that this be recognized before the city is included in the test of the accuracy of such an estimating technique.

Also needed is a series of tests of the reliability of school enrollment as a measure of population of a specific age group, rather than the present implicit assumption in many tests that differences from a census standard

result only from the operation of the estimating method. This is important because errors in the measurement of the population of school age become magnified in the natural-increase-and-net-migration methods.

The use of special censuses by California cities has provided a basis for testing of estimating methods against census standards, but the distribution of special census dates through time operates to limit the strict comparability of results. For most tests, therefore, a distinction is made between estimates for dates in 1955 and later, as compared with tabulations that include estimates for earlier years.

An initial test was made of the Census Bureau's method II as applied to 12 California cities, the total number of cities for which necessary data by age has been published. These estimates differed on the average by 5.8 percent from the published census figures. It was possible to apply method I to a total of 26 cities that had taken 34 special censuses after April 1, 1950. In using this method, an adjustment factor was incorporated into the procedure on the basis of earlier experimental work which showed that greater accuracy should result. For the total of 34 special censuses, the average difference of the estimates was 3.6 percent, while estimates for the 17 special census dates after January 1, 1955 differed from the census results by an average of 4.9 percent. Using the same 12 cities included in the method II test, the average deviation for method I was 4.3 percent. When the method I and II estimates were averaged for the 12 cities, the average difference was reduced to 3.4 percent.

A second test combined a partial-direct variable, live births, and a total-direct variable, deaths, with a method that measures total population change, that is, Bogue's vital rates method. Forty-five cities were included in this test, and the average difference of the estimates was 11.9 percent. Only 13 of the 45 estimates were within 5 percent of the special census results and 23, or more than half, differed by 10 percent or over. As might be expected, the estimates were predominantly lower than the special census results.

It is evident that the vital rates method is at best applicable only to large and very slow-growing cities. This is confirmed by data in Shryock's paper on the "Development of Postcensal Population Estimates for Small Areas", presented at the 1957 Conference on Research in Income and Wealth, showing that the "natural increase method", which makes use of birth and death statistics and assumes that net migration equals zero, showed smaller average deviations than the vital rates method for the 92 cities included in the test.

The third class of estimating technique was termed the composite, because it combines both component-of-change and total-change methods. Some experimental work has been carried out with the age-specific-death-rate method, which may be used to estimate the population aged 15 years and over in combination with either method I or II for estimating the population under 15 years of age. Estimates for cities were prepared, using deaths by sex and by 10-year age groups. For eight cities included in the initial test, estimates of the population aged 15 years and over differed from special census totals by an average of 4.7 percent. Interestingly enough, the greatest differences in age groups appeared in the 15-24 year, 25-34 year and 75 year and over age groups.

In 1954 a test of the age-specific-death-rate method was made, deriving 1950 population estimates for 29 states with relatively small nonwhite populations. These states were used so that 1940-50 change in the age-specific white death rates for the United States could be used in estimating changes in state death rates. The estimates of total population aged 15 years and over differed from the 1950 census by an average of 4.6 percent. For the population aged 15 years and over, the average deviation for the 29 states was 2.4 percent. These results point to the utility of the age-specific-death-rate method for preparing estimates by age, particularly for the older age groups. In this same test, a check of the 1950 population estimates by 10-year age groups for ages 15 through 84 years and 85 years and over for the five largest states, New York, California, Pennsylvania, Illinois and Ohio, showed that the estimate for only one age group -- the 15 to 24 year olds in Pennsylvania -- differed by more than 5 percent from the census figure.

The discussion so far has dealt only with partial-direct and total-direct indicators, involving the use of symptomatic data expressed in numbers of people. The tests of methods and indicators suggested the need for further efforts to improve the accuracy of city estimates. The next step was to examine the availability, accuracy, and applicability of such indirect indicators as building permit records and utility data.

One test endeavored to explore the possibility of estimating change in the number of persons per household, assuming that an accurate measure of change in the number of households could be achieved. That is, the reported number of occupied dwelling units at the time of the special census was used to measure change in dwelling units from 1950, while change in the number of persons per household was estimated by relating each city's average to the national pattern as reported in the Current Population Survey. Population estimates were prepared for 59 California cities, including 66 special censuses in 1955 and 1956. The average deviation of the estimates was 3.63 percent. These cities ranged in size from 1,700 to 2,244,000 at the time of the special censuses, with growth after the 1950 Census ranging up to 536 percent, and with at least 10 cities reporting over 100 percent gains.

A second test dealt with the use of utility records as a basis for estimating change in the number of occupied dwelling units or households. Through the cooperation of the major utilities in the State, information for almost all of the more than 360 cities in California is available for use in tests and estimate preparation. The major problems of the use of electric meters for this purpose are well known: the existence of master meters particularly in public housing and in trailer parks; the tendency to leave meters active although the premises are vacant; the effects of economic change on the relationship of meters to households; and others. In tests to date, estimates of the number of occupied dwelling units have been prepared for 44 cities that have taken 64 special censuses since January 1, 1955. In a test of 26 special censuses taken in the San Francisco Bay Area, the average deviation of the estimates of the number of occupied dwelling units was 2.5 percent from the reported figures, with 5 cases showing differences of 5 percent or more. These five cities are all located in one county and involve areas where war housing and master metering was operative in 1950, with considerable change since then. A check of estimates applied against 38 special censuses in the Los Angeles-Long Beach Metropolitan Area showed an average deviation of 1.5 percent from the reported occupied dwelling units, with the largest deviation under 5 percent. In the Southern California test, the benchmark dates were special censuses in 1952 and 1953 rather than the decennial census of 1950.

These data suggest that it is possible to prepare estimates of the numbers of dwelling units in cities that are sufficiently accurate to stand as estimates of the number of households for those agencies needing such figures. Further, it would appear that satisfactory estimates of population can be derived from these data. One test, applied to 23 cities in the San Francisco Bay Area, related percentage change in the number of occupied dwelling units based on utility data to percentage change in the total population, using an equation based on census data from 32 cities that took special censuses between January and April of 1956. The average difference of the 47 estimates for these 23 cities from the census standards was 4.2 percent, with 6 estimates differing more than 10 percent.

The Census Bureau, in its use of this general type of approach, has included a modification in which the numbers of persons per household under 18 years of age and 18 years and over have been handled separately. This modification has been tested, using the national pattern of change as a guide for estimated local averages. For the same 12 cities used in the test of methods I and II, estimates of total population aged 18 years of age and over showed an average deviation of 2.5 percent from the special censuses. At the present time this procedure is used for all official city estimates where a basis exists for establishing trends of change for the two age groups.

Another modification has been incorporated in the preparation of recent estimates for the cities of Los Angeles and Oakland. With the cooperation of the City Planning Commission staffs, data were developed for four subareas in Los Angeles and five subareas in Oakland, so that allowance could be made for subarea differences in change in the number of persons per household. It is believed that both estimates gained greater accuracy through this intermediate step, because both cities display markedly different patterns between subareas.

While the conclusion must be considered tentative, it is important to observe that the tests of the dwelling unit method suggest that neither the size of a city nor its rate of growth appears to influence the accuracy of an estimate. On the other hand, only the taking of a large number of special censuses has provided assurance that change in the number of persons per household can be traced over the years since the last decennial census. Further, each estimate must be developed individually in order to achieve maximum accuracy.

Do we have effective means for measuring population change in small areas? The answer would appear to be a qualified "yes". A variety of methods exist for translating symptomatic data into population measures, at least for the most widely recognized formally defined small areas such as states, counties and incorporated places. With varying degrees of applicability and accuracy, these methods will describe total populations, characteristics such as age, sex and color, components of change, numbers of households, and so forth. The degree of effectiveness is likely to depend in large measure on the recognition that indicators and methods must be related to specific cases. This is illustrated by the fact that limited testing of method I on California cities led to the introduction of a relatively simple adjustment factor that results in method I giving generally more accurate estimates for California cities than method II. It is quite possible that a further analysis of method II as applied to cities will again reverse the relationship, although there is no inherent reason for one method being more accurate than the other. Their differences arise primarily from the age groups selected as partial-direct indicators of migration.

The comments of the Census Bureau on the very limited extent to which state and city agencies report testing of the methods they use points to the need for a considerable expansion of testing activity before population estimates are accepted as satisfactory in terms of the criteria applicable. Admittedly this type of testing cannot always be done, but it may be that the producers of population estimates must share some blame for this in that they have not convinced the users -- the persons who originate requests and utilize results -- that the lack of sufficient time and adequate facilities to collect and check symptomatic data and to test and modify methods can only reduce the effectiveness of the population estimates and limit their value as administrative tools.