SOME EFFECTS OF STATISTICAL SYNTHETIC ESTIMATION ON CENSUS UNDERCOUNT OF SMALL AREAS

Linda K. Schultz, Elizabeth T. Huang, Gregg J. Diffendal and Cary T. Isaki, U.S. Bureau of the Census

I. Introduction

Currently the Bureau of the Census is examining different methods of adjusting the decennial census for undercount. Besides examining methods by which the census might be adjusted, the Bureau has been acquiring actual operational experience on the issue of adjustment with the 1986 Census test currently being conducted in a section of Los Angeles, California. This test is to help determine whether an adjustment is operationally feasible in the time frame specified. The work to be described below does not discuss operational issues but examines a proposed adjustment method termed statistical synthetic. This paper will examine how well two statistical synthetic estimators and the census perform at the state and county levels of aggregation using several measures of improvement. Known adjustment factors will be compared with estimated factors as to their effects on the state and county population estimates. Estimates of population at the enumeration district level will also be examined.

II. Background

In recent years there has been much discussion and interest in the issue of census adjustment. The Bureau of the Census has been investigating different adjustment methodologies as part of its ongoing research effort. Some examples of the methodologies can be found in Cowan and Fay (1984), Diffendal, Isaki and Malec (1982), Diffendal, Isaki and Schultz (1984), Isaki, Schultz, Smith and Diffendal (1985) and Isaki, Diffendal and Schultz (1986). The purpose of this paper is to discuss one of the adjustment methodologies, termed statistical synthetic, under investigation. In this procedure survey based estimates of total population by strata are used to construct adjustment factors. The strata are formed by grouping persons believed to possess similar undercount rates. For example, one stratum could consist of all Black males in the cities of Chicago and Detroit. In the statistical synthetic procedure, the adjustment factors are applied to census block counts of the appropriate stratum. Block counts are the basic building block of the census. Hence, when the blocks are adjusted, all levels of aggregation result in arithmetically consistent tabulations.

As with any adjustment methodology we would like to evaluate how well the adjustment compares with the true population counts. Unfortunately this is not very easy. We do have two other sources of estimates of the population count besides the census, demographic analysis and the post enumeration program (PEP) for 1980. However, both demographic analysis and the PEP have disadvantages as tools for evaluating small area estimates.

The demographic analysis program provides estimates of the legal population by age-race (Black, non-Black)-sex at the U.S. level. Because the estimates only exist at the U.S. level it is difficult to evaluate the accuracy of adjustments for states and subunits within states. The second source of population estimates is the PEP (Cowan and Bettin (1982)). While the PEP is the only source of estimates of undercount at the state and large metropolitan area level, the estimates do have a number of deficiencies as outlined by Freedman and Navidi (1984). As with demographic analysis, PEP estimates at lower levels, in general, do not exist. Since neither of the latter two sources of undercount estimates are satisfactory for evaluating small area adjustment methods below the state level we constructed several data sets of artificial population counts.

A. Artificial Populations

The construction of the artificial populations, termed AP2 and AP3 (to be used below), is described by Isaki, Diffendal and Schultz (1986). Briefly speaking, adjusted census substitution figures were used as a proxy for undercount. Census substitutions are the result of imputing people into housing units. The differences in AP2 and AP3 consisted in how the substitution figures were adjusted to form the artificial undercount. AP2 assumes Hispanics are similar to the non-Black population while AP3 assumes that Hispanics and Blacks are similar. The artificial undercounts were then added to the census counts to form the artificial population totals.

B. Estimation

Using the two artificial populations described above we examined two estimators, syn 2 and syn DA. (A third estimator, syn 1 described in Isaki, et.al. (1986) was omitted in the study as it was basically inferior to the remaining estimators).

The synthetic estimation strategy for adjustment of census undercount in small areas is based on the construction of strata of persons felt to possess similar undercount rates. The adjustment factors for syn 2 were defined according to our perception as to how undercount (artificial population minus census count) might vary across strata. The factors were defined without knowledge of the distribution of the undercount in regard to either of the artificial populations. The main interest was in how the estimators performed in estimating small area population counts of the artificial populations in the absence of sampling error.

The adjustment strata for syn 2 were based on census divisions, as well as size of place and race. For example, within a census division all Blacks in cities with more than 250,000 persons might be considered as one adjustment strata. Syn DA's adjustment factors were defined by age-race-sex at the U.S. level only.

In general, a synthetic estimator $\hat{T}_i$ for area i has the form

$$\hat{T}_i = \sum_j F_j C_{ji}$$

where C_{ji} is the census count of persons in stratum j in area i and F_j is the adjustment factor (ratio of the total artificial population count for stratum j to the census count for stratum j computed over a collection of areas including area i).

III. Results

We produced the synthetic estimate for each artificial population at the enumeration district, ED, level and placed the results on a file together with comparable census and artificial population counts by total population and individual race groups. The data were then tabulated to larger geographic levels and the adjusted figures compared to the census. We then evaluated the adjusted figures at different levels of geography such as state and county. We also have measures of improvement based on ED's for two states. In the next section the measures of improvement will be described. To assess the performance of the estimation strategies we chose summary statistics related to population counts. We also examined the effects of each estimation method on apportionment.

A. Summary Statistics

The measures of improvement or summary statistics used for examining states and counties can be dichotomized into those dealing with absolute relative errors (ARE) and those dealing with absolute differences in proportions (ADP). The ARE treats estimation error of each area relative to its population size. The ADP summarizes the performance of the estimation strategies in providing correct proportions of the population rather than their correct level.

In defining such statistics it is necessary to introduce some notation. Let subscript i denote the i-th area of interest and L denote the number of areas i under consideration. Also, define AP_i to be the artificial population count, C_i to be the census count and E_i to be the synthetic estimate for area i. Define and denote the absolute relative errors of the synthetic estimate for area i by

$$ARE(E_i) = (AP_i)^{-1} \mid AP_i - E_i \mid .^*$$

The mean absolute relative error is defined as follows:

$$MARE(E) = L^{-1} \sum_{i=1}^{L} ARE(E_i). ^*$$

Given the manner in which the artificial populations were constructed the MARE of the census was expected to exceed that of the synthetic estimators. Another ARE measure examined in this work was

$$MSRE = L^{-1} \sum_{i=1}^{L} AP_i \, (ARE\,(E_i))^2. ^*$$

In describing the ADP measure some notation is necessary. Let the i-th area's proportion of artificial population counts, census counts and synthetic estimates be defined and denoted by

$$P_i^A = (\sum_{i=1}^{L} AP_i)^{-1} AP_i, \quad P_i^C = (\sum_{i=1}^{L} C_i)^{-1} C_i$$

and $P_i^E = (\sum_{i=1}^{L} E_i)^{-1} E_i$, respectively. Furthermore, let the absolute difference of proportions for the synthetic estimator for the i-th area be defined and denoted by

$$ADP(E_i) = |P_i^E - P_i^A|. ^*$$

The first measure under ADP, is the sum of absolute differences of proportions, SADP, of the estimator and the artificial population:

$$SADP = \sum_{i=1}^{L} |P_i^E - P_i^A|.^*$$

Values of SADP smaller than that of the census favor the synthetic estimator.

Another measure, proportion of population improved, PI, is defined by

$$PI = [\sum_{i=1}^{L} AP_i]^{-1} \sum_{i=1}^{L} IMPV_i \quad \text{where}$$

$$IMPV_i = \begin{cases} AP_i & \text{if } ADP(E_i) < ADP(C_i) \\ 0 & \text{otherwise} \end{cases} .$$

PI indicates the proportion of the total artificial population which resides in areas whose proportions of the total artificial population after adjustment are closer to their true proportions than are their proportions before adjustment. The larger PI exceeds .5, the more favorable is the adjustment in terms of population affected. The measure PI was used by Schirm and Preston (1984).

The measure of misproportionality (MISPROP) was also examined. This measure sums the squared differences in relative errors of the adjusted numbers within the nation or a state with the relative error of the nation or state respectively, weighted by the artificial population. The measure can be defined as follows:

$$MISPROP\ (E) = \sum_{1}^{L} AP_i \times D_i \ ^*$$

$$\text{where} \quad D_i = [(E_i - AP_i)\, AP_i^{-1} - (E_T - AP_T)\, AP_T^{-1}]^2$$

$$\text{and} \quad E_T = \sum_{1}^{L} E_i, \quad AP_T = \sum_{1}^{L} AP_i$$

B. States

In this section, comparisons are made of the synthetic methods for two artificial populations by total population for the 50 states and DC (L = 51).

For the artificial populations, the synthetic estimates exhibited smaller error than the census for all measures of improvement examined. Syn 2 almost always indicated smaller error than syn DA.

The improvement of syn 2 over syn DA is especially noticeable with the sum of absolute differences in proportion (SADP) measure. While the syn 2 and syn DA SADP measures are smaller than the census SADP measure, syn 2 is the only one with a noticeable improvement over the census. With regard to the census for both ARE and ADP measures, no synthetic strategy produced state adjustments that were superior for every state. According to the PI measure however, the proportions of the population in states where adjustment improved on the census ranged from .62 to .87. The measure of misproportionality also showed syn 2 as being superior to both syn DA and the census for both AP2 and AP3. The measure indicated that syn DA, while not quite as good as syn 2, was much better than the census.

The apportionment issue was also examined. We compared the apportionment that would occur based on the 1980 Census with the artificial population and the adjusted populations. The tables below show the states that are affected by the adjustments.

After examining the table one can observe that syn 2 is more consistent with the artificial population counts than the census or syn DA in regard to apportionment.

State whose seats are affected	Seats in the census	Change from census according to - AP	Syn 2	Syn DA
	Apportionment Application - AP2			
1. Alabama	7	+1	+1	+1
2. Georgia	10	+1	+1	+1
3. Indiana	11	-1	-1	-1
4. Iowa	6	-1	-1	
5. Kansas	5	-1	-1	-1
6. New York	33	+1	+1	
	Apportionment Application - AP3			
1. Alabama	7	+1		
2. Georgia	10	+1	+1	+1
3. Indiana	11	-1	-1	-1
4. Iowa	6	-1	-1	
5. Kansas	5	-1	-1	-1
6. New York	33	+1	+1	+1
7. Ohio	21	-1		
8. Texas	27	+1	+1	

C. Counties

At the county level we compared the adjustment methods to the census for total population. For AP2 and AP3, both of the adjustment methods were better than the census for the measures considered.

Unlike the state analysis we found syn DA to be superior to syn 2 for both AP2 and AP3. Syn DA had the smallest MARE and the smallest median ARE. This result was consistent for both AP2 and AP3. Syn 2's SADP measure was slightly lower than the SADP measure for syn DA. However, when examining the number of counties in which the census had a lower ADP than the adjustment we find that the adjustment methods yielded similar results. While syn 2 had a lower SADP measure the percent of counties in which the census was superior to syn 2 was not much different from the results for syn DA. Likewise the PI measures did not differ much between the adjustment methods.

We also divided the counties by population size into three groups and analyzed each group separately. The three groups were counties with population 0 to 10,000; 10,000 to 50,000; and those with populations greater than 50,000 with approximately 25%, 50% and 25% of the counties in each group, respectively. This analysis indicated that syn DA appeared to fare the best for counties with smaller population sizes while syn 2 did well for counties with larger population. For counties in the middle range (10,000-50,000) the results were not clear, for some measures syn 2 appeared to be the best and for others syn DA looked better.

In further analysis we examined the counties within a given state, state by state. In examining the median ARE measure for each state we found that the magnitude of the median ARE was about the same order of magnitude as the median ARE for all 3137 counties. Unlike the overall state results, the median ARE for the census was not consistently worse than the median ARE for the adjustment methods. In fact for AP2 there were 11 states in which the census had a lower median ARE than the adjusted counts by either adjustment method and for AP3 there were 17. While there are states in which the census had the lower median ARE, the states, in general, represented the less populated areas. Therefore, for a majority of the country's population, adjustment continues to be superior to the census. Similar results were found in the examination of the MSRE measure for both AP2 and AP3.

We also examined the measures of misproportionality and proportion improved for counties within a given state, state by state. States with relatively high misproportionality included California, Florida, Georgia, Illinois, Louisiana, New Jersey, North Carolina, Tennessee and Texas for AP2. For AP3 the previous list of states plus New Mexico were the states with high misproportionality measures. New York and Pennsylvania were unique compared to the other states because they both had high measures of misproportionality for the census indicating misproportionality while after adjusting the census counts by either adjustment method the measure of misproportionality is considerably reduced. This holds for both artificial populations. In general the measure of misproportionality is consistent over the two artificial populations. It was only with New Mexico and Texas that the artificial populations differed substantially. This may be attributed to the assumptions made in constructing the artificial populations.

The proportion improved measure shows that in 66%-70% of the states syn DA is closer to the artificial populations than the census for at least half of the total population within the given state. For syn 2 72%-74% of the states are improved for at least half of the population within a given state.

D. Enumeration Districts

North Dakota and Mississippi were examined at the ED level using the measures of improvement. North Dakota was chosen to represent states with smaller total population. Mississippi was selected as a state of more moderate size.

North Dakota had 2536 EDs that were examined. In our analysis using AP2, we found that the MARE was .0016 for the census, for the two estimators the MARE measures were between 2 and 3 times larger. The MARE for syn DA was .00363, for syn 2 the MARE was .00423. Using AP3 we found that the MARE measures doubled when we went from the census to the estimators.

A similar trend was noted when the 3595 EDs of Mississippi were examined. Again the census MARE measure was notably smaller than the MARE measures for syn DA and syn 2. This occurred for both AP2 and AP3.

The results based on the ARE measures at the ED level for North Dakota and Mississippi contrast sharply with the state and county results. The results indicate we may have a problem using either syn DA or syn 2 at lower levels of geography for states similar to North Dakota and Mississippi.

IV. Sampling Error

In the results presented above we have shown that it is possible, using statistical synthetic estimation and a number of assumptions, to improve on the census at the state and county levels of geography. We have computed adjustment factors based on all the enumeration districts (ED's) in the country. However, in practice if we were to implement this procedure the adjustment factors would be computed based on information gathered from a sample, meaning the adjustment factors would be estimated from partial knowledge rather than full knowledge. So how does this affect the results presented above?

Using the ED files we did a simulation that can be used to investigate this question. The adjustment factors were computed based on a sample of ED's. We used the two estimators, syn 2 and syn DA, and the artificial populations, AP2 and AP3. Since it is projected that a block sample will be used in a post enumeration survey for 1990, and we only have ED information, we attempt to get a rough idea of what will happen using a block sample by using a sample size of 1442 ED's to get a bound on the magnitude of possible errors. It should be stressed that in the work that follows, we assume that the adjustment factors are estimated unbiasedly.

We selected 90 samples using systematic sampling of ED's within strata within each of the nine census divisions. Each of the 90 samples generated a set of adjustment factor groups. The number of adjustment factors in each group depends on whether the estimator is syn 2 or syn DA. For syn 2 there will be 96 factors; for syn DA there will be 30. Ideally, based on a given sample size, estimator and artificial population we would like to use the estimated factors to generate 90 estimates of population for each state and county for the total, Black, Hispanic and Rest populations and include the results on a file so that we could also examine the distributions of the summary measures. This, however, given the cost of computing, was not possible. A random selection of one of the 90 replicates was made and the resulting adjustment was based on the adjustment factors estimated from the single replicate. The results are presented in the next section.

V. Sampling Results

In this section we present the results found when we analyzed one of the 90 replicates and its resulting adjustment to state and county total population estimates. We examined both syn 2 and syn DA when the adjustment factors were formed based on a sample of ED's rather than a tabulation of all of the ED's. The results were consistent over both artificial populations. While the sampling had no apparent effect on syn DA, the same cannot be said for syn 2. For AP2 at the state level, syn 2's MARE measure increased from .0044 before sampling to .0060 after sampling. Syn 2's PI, proportion improved measure, dropped from 70% to 48%. A PI measure below 50% indicates that the adjustment did not generate an improvement. Similar results, although not as severe, occurred for AP3 at the state level. The MARE measure increased from .0045 to .0060, and the PI measure dropped from .872 to .635. In the case of AP3 an improvement over the census is still occurring for syn 2. Syn DA's PI measure of 70% lends support to the observation that syn DA appears to be superior to syn 2 when sampling error occurs. Results at the county level further support these observations. When sampling error is included in the adjustment factors syn 2 does not perform as well, possibly because we are forming 96 adjustment factors versus only 30 for syn DA.

Apportionment results do change when sampling error is introduced. Syn 2 no longer agrees exactly with AP2 as it did when there was no sampling error. Alabama would have 1 too few representatives and Iowa 1 too many using AP2 and the syn 2 adjustment method. Using the syn DA method Alabama would have the correct number of representatives, Iowa would have 1 too many and New York 1 too few. Looking at the apportionment results for AP3 we find the two adjustment methods agree with each other. When compared to AP3 Alabama is too low, California too high, Ohio is also too high and Texas is too low. The tables below show the states that are affected by the adjustments in the sense that the seats allocated to them would change from the census.

State whose seats are affected	Seats in the census	Change from census according to - AP	Syn 2	Syn DA
	Apportionment Application - AP2			
1. Alabama	7	+1		+1
2. Georgia	10	+1	+1	+1
3. Indiana	11	-1	-1	-1
4. Iowa	6	-1		
5. Kansas	5	-1	-1	-1
6. New York	33	+1	+1	
	Apportionment Application - AP3			
1. Alabama	7	+1		
2. California	45		+1	+1
3. Georgia	10	+1	+1	+1
4. Indiana	11	-1	-1	-1
5. Iowa	6	-1	-1	-1
6. Kansas	5	-1	-1	-1
7. New York	33	+1	+1	+1
8. Ohio	21	-1		
9. Texas	27	+1		

VI. Conclusions and Future Work

There are two important facts to keep in mind before drawing any conclusions from the work presented in this paper and its applicability for census adjustment. First of all the results are highly dependent on the assumption that census substitutions are a reasonable proxy for undercount. All of our results are relative to the artificial populations we have formed using that assumption. If this assumption is not reasonable our results may produce an inaccurate picture. Second, even if substitutions are adequate as a proxy for undercount we have only looked at two possible statistical synthetic estimators.

The state results indicate that syn 2 is superior to the census as well as to the other adjustment methods. Syn 2 outperformed syn DA for total population for most of the measures of improvement. Syn DA was slightly better than the census.

The county total population results suggest that syn DA is superior to syn 2 as well as being superior to the census. However, when we divided the counties into groups by their sizes the results were different. While syn DA was superior to the other adjustment methods for small counties, syn 2 did better for the larger counties. This result suggests there may not be one synthetic adjustment that is satisfactory for all areas but that we may need to apply separate strategies over portions of the universe of areas.

The examination of counties within states illustrated that the success of any particular adjustment method is not going to be clear cut. By partitioning the counties by state we found that for some states the unadjusted census would be the preferred method while for others one of the adjustments would be better. While the PI measure indicates that the adjustments improve a majority of the counties within a majority of the states, other measures indicate the improvements generated may not yield substantial gains over the census.

The introduction of sampling error into the adjustment factors illustrated that syn 2 is affected much more than syn DA. In fact sampling error causes syn 2 to lose out to syn DA as the preferred method of adjustment at the state level.

Another important issue is whether the results, for example, at the state level are good enough. We've shown we can improve on the census at a higher level of geography such as states, but the adjustment methods discussed here appear to falter at lower levels of geography.

References

1. Cowan, Charles D. and Bettin, Paul J. (1982). "Estimates and Missing Data Problems in the Postenumeration Program," technical report, U.S. Bureau of the Census, Washington, D.C.

2. Cowan, Charles D. and Fay, Robert E. (1984). "Estimates of Undercount in the 1980 Census," Proceedings of the Survey Research Methods Section of the American Statistical Association, Washington, D.C.

3. Diffendal, Gregg J., Isaki, Cary T. and Malec, Donald J. (1982). "Examples of Some Adjustment Methodologies Applied to the 1980 Census," technical report, U.S. Bureau of the Census, Washington, D.C.

4. Diffendal, Gregg, Isaki, Cary and Schultz, Linda (1984). "Small Area Adjustment Methods for Census Undercount," invited paper to the Data Users Conference on Small Area Statistics, U.S. Department of Human Services, Washington, D.C. 52-56.

5. Ericksen, Eugene P. and Kadane, Joseph B. (1985). "Estimates of the Population in a Census Year - 1980 and Beyond," Journal of the American Statistical Association, 80, 98-109.

6. Freedman, D.A. and Navidi, W.C. (1984). "Regression Models for Adjusting the 1980 Census," Technical Report No. 35, University of California at Berkeley.

7. Isaki, Cary T., Schultz, Linda K., Smith, Philip J. and Diffendal, Gregg J. (1985). "Small Area Estimation Research for Census Undercount — Progress Report," Statistical Research Division Report Series, Census/SRD/RR-85-07, Bureau of the Census.

8. Isaki, C.T., Diffendal, G.J. and Schultz, L.K. (1986). "Statistical Synthetic Estimates of Undercount for Small Areas," paper presented at the Census Bureau's Second Annual Research Conference.

9. Schirm, Allen L. and Preston, Samuel H. (1984). "Census Undercount, Adjustment and the Quality of Geographic Population Distributions," technical report, University of Pennsylvania.

10. Tukey, John W. (1984). "Points to be Made," presented to the committee on National Statistics, Panel on Decennial Census Methodology, May 11, 1984.

Footnote

*This measure can be computed for the census by substituting C_i for E_i.