BAYES ESTIMATES OF POPULATION UNDERCOUNT FOR LOCAL AREAS

Eugene P. Ericksen, Temple University
Department of Sociology, Philadelphia, PA 19122

I. THE PROBLEM

In 1980, the Census Bureau spent $1.1 billion on the census, more than four times the actual cost and twice the inflation-adjusted cost of the 1970 Census. The policy of the Bureau was to make every attempt to attain complete coverage, and several coverage improvement programs were included in the census plan. Some of these, like hiring additional staff in district offices, were intended to make the whole census operation better. Others, like the Vacant-Delete Check or the Non-household Sources Program, were focused on specific counting problems.

Comparing 1970 and 1980, the Bureau did appear to improve the count. The net undercount, or difference between omissions and erroneous enumerations, dropped from 2.9 percent in 1970 to 1.4 percent in 1980. The Black undercount also dropped from 8.0 percent in 1970 to 5.9 percent in 1980. Even so, serious questions remained about the accuracy of the count. If the Bureau had reduced omissions, or people improperly excluded from the count, the undercount would have been reduced and the census improved. On the other hand, if there was no change in omissions but erroneous enumerations, usually duplications or fabrications, had increased, the net undercount would have been reduced but the census made worse.

According to a Census Bureau report (1988, ch. 3), much of the apparent improvement was due to an increase in duplications, thought by the Bureau to be few in number in 1970. If we use the Bureau's best estimate of 2.5 million duplicates, the improvement from 1970 to 1980 was only 600,000 people. If we add the 2.5 to 3.5 million erroneous enumerations other than duplicates, the improvement falls to zero or perhaps becomes negative. Evaluation data from the 1970 and 1980 censuses indicate that the omissions rate at least remained constant but probably increased between 1970 and 1980 (U.S. Bureau of the Census, 1975, 1982). At best, by spending more money in 1980, the Census Bureau slightly reduced the undercount problem. More likely, it created the illusion of improvement while the problem in fact got worse.

In any event, a substantial undercount differential between Blacks and non-Blacks remained. In 1980, the non-Black rate was 0.7 percent, 5.2 percentage points lower than the Black rate. In 1970, the corresponding differential was a slightly higher 5.8 percent. In its report, the Bureau states that it reduced the undercount more in the South than in other areas. Because census-taking problems appear to be worse in large cities, it is likely that Blacks and other minorities were harder to count there. Many large northern and western cities have substantial minority populations, and therefore probably had much larger undercounts than the rest of the country. This evidence of a differential undercount creates political pressure for adjustment.

The problem is that the data which might be used to adjust the count are imperfect. Demographic analysis provides national, but not local, estimates. While Schirm and Preston (1987) have demonstrated that synthetic estimates based upon applying national race-specific undercount rates to local areas are likely to improve upon the raw count, such synthetic adjustments would not alter the figures very much. Because it is harder to count Blacks and other minorities living in poor neighborhoods of large cities, the synthetic adjustment is not likely to go far enough.

Survey data from the Census Bureau's Post Enumeration Program (PEP) are a better data source. The sample sizes were large enough that separate estimates of the net undercount were produced for the 50 states, 23 metropolitan areas, and 16 central cities. The difficulty is that, for an important part of the sample, the Bureau could not tell if people were counted or missed (Cowan and Bettin, 1982). The omissions rate estimates were based on the April and August editions of the Current Population Survey. These samples were matched against the census to estimate the proportions of people who were omitted. In some cases, the CPS address could not be found in the census; in other cases, either the CPS or the census data were so poor that no one could tell if there was a match. For about four percent of cases, there was no conclusive determination, and imputation was required. About one-third of cases ultimately classified as "omitted" had this status assigned by imputation. The problem is serious because the sorts of people less likely to be counted in the census are also less likely to provide good data to the CPS interviewer. Characteristics data such as race and the nature of household attachment which might be used to find a donor for imputation are measured inexactly. Such people are also less likely to be around when the follow-up interviewers from the Census Bureau go to the field to try to resolve discrepancies between the Census and the CPS.

This creates a tough choice. We are confident that an undercount exists, and it is almost certainly differential. At the same time, the data which we might use to estimate these differentials are imperfect. How might we evaluate all the data to come to an appropriate statistical policy? In this paper, I will argue that we must evaluate the likely nature of error in both the Census and evaluation data to develop a good policy. In the sections below, I will briefly describe how the census works. Next, I will describe how omissions and erroneous enumerations occur and present evidence of their distributions. Finally, I will describe how statistical models might be used in making the decision.

II. HOW THE CENSUS WORKS

Given the cost, the Decennial Census is obviously a complex process. To think intelligently about adjustment, we need to understand how the Census works, and I attempt a short description in this section. There were basically three steps.

The first step was to compile the Master Address Register (MAR), done in essentially two ways. In urban areas, the Bureau obtained commercial address lists and updated them by using information from the Post Office and by sending field-checkers out to the local areas. In less urban areas where commercial lists were not available, census enumerators compiled the initial list by observation, and this was checked for completeness by later field-checking. The MAR's were used in 95 percent of the country. In the remaining five percent, generally sparsely populated rural areas located west of the Mississippi River, no Master

Address Register was compiled, but area maps were prepared for the enumeration.

The second step was to mail the census forms out and have households mail them back. In 1980, 83 percent of households did so, but these rates varied by area. In many suburban and other less urban areas, mailback rates were well above 90 percent, but in poor minority neighborhoods of cities like Chicago and New York, rates were often below 70 percent, and sometimes well below. In conventional areas, there was no mail-out of census forms. Instead, enumerators went to the field with their maps, listing addresses and counting people through personal interviews.

The third step was to "follow-up" at those households who did not mail their forms back in. This is a difficult part of census-taking, as many people are hard to find at home, good enumerators are scarce in many areas and some neighborhoods are dangerous. Because of the variation in mailback rates, the volume of work was much greater in some neighborhoods than others, e.g., in a neighborhood with a mailback rate of 60 percent, there was four times the amount of follow-up work as in a neighborhood where the rate was 90 percent.

Census-takers were told to make several attempts to contact and enumerate each household. If, after several attempts, no contact was made, the "close-out" procedure was used with information obtained from neighbors or passersby. In the extreme case, when no one was found to provide the information, the occupancy/vacancy status of a housing unit was imputed by computer and if deemed to be occupied, the number and characteristics of the occupants were also imputed by computer. Combining "close-out" and "no information" cases, a total of 3.3 million people were imputed into the census, 1.47 percent of the total.

The Bureau, at a cost of $69.75 million, used a number of coverage improvement programs in 1980, increasing the count by 2.6 million people. The majority of these were added by the Vacant-Delete Check, in which every housing unit initially listed as vacant was re-checked to determine whether it was really vacant. The Bureau estimates that 1.7 million people were added by this procedure. The other coverage improvement programs were less successful, the difficulties often due to the lack of well-trained census personnel to carry the programs out properly (General Accounting Office, 1980; U.S. Bureau of the Census, 1987).

III. OMISSIONS

The Census Bureau (1988) estimates the net undercount of the 1980 Census to be 3,171,000. When we add the 3.3 million imputations and five to six million erroneous enumerations to this figure, we estimate the total number of omissions to be between 11.5 and 12.5 million. Why did they occur in such large numbers?

Missed housing units are part of the problem. According to a Census Bureau study (Fein and West, 1988), 39 percent of 1980 Census omissions occurred because the entire housing unit was missed and probably left off the Master Address Register. Such omissions are likely to be concentrated in cities and rural areas. In cities, housing units are frequently hard to find in situations where a single family house has been broken into apartments, but only a single door, bell, and mailbox are visible, or where a primary family rents out part of its house. While the structure is visible, the census-taker does not identify and list all housing units. In rural areas, housing such as that for migrant workers is often not visible from a road. Moreover, census-takers sometimes have difficulties reading maps and do not canvas their areas completely. If the housing unit is not included in the Master Address Register, it will not be sent a form, and it is much less likely that it will be included in the census. According to a survey of New York City residents concerning their participation in the 1980 census, only 32 percent of persons who said they did not receive a census form could remember being contacted or visited by a census-taker (Schoen, 1980).

The Census Bureau reports that the majority of missed persons live in housing units where others are counted. Recent studies (Fein and West, 1988; Boone, 1987) have focused on the question of who gets missed within households. Each study relied on a survey of missed people living in urban minority neighborhoods. Missed people were more likely to live in large households and were less likely to be in the nuclear family of the household head. For example, cousins, grandchildren and unrelated boarders were more likely to be missed. Recent immigrants, persons not speaking English well and welfare recipients were also hard to count. There was a small amount of purposeful concealment. The main result is that people living in small nuclear families are more likely to be counted than are people whose family attachments are less strong. Unconventional living arrangements are more frequent in poor neighborhoods of large cities.

This combination of factors has caused omission rates to be much higher in large cities than elsewhere. According to the PEP Series 3, usually relied upon by the Census Bureau in its research on the undercount, the omissions rate for 16 large cities (e.g., New York, Los Angeles and Chicago) was 9.7 percent. In 20 states and state remainders (excluding the 16 cities), where at least 10 percent of the population was Black or Hispanic, the omissions rate was 6.2 percent. In the predominantly White remainder of the country, the omissions rate was 3.5 percent. The chances of omission are clearly greater in large cities than elsewhere. A National Academy of Science panel analyzed the data another way. They found omissions rates to be highest among Blacks and Hispanics living in areas where over 35 percent of people failed to mail their forms back in (National Research Council, 1985, p. 227). Such areas are almost exclusively found in central cities.

IV. ERRONEOUS ENUMERATIONS

While omissions are concentrated in central cities, it is logically possible that they would be balanced by erroneous enumerations. There were approximately 6.5 to 7.5 million of these in the 1980 Census (Cowan and Fay, 1984; Fay, 1988a). Some of these, about 1.5 million, were geocoding errors in which a person was counted only once, but in the wrong place. In other words, an address was assigned to the wrong block, census tract or enumeration district. The remaining 5 to 6 million errors were "people" who should not have been included in the count at the address where they were listed. These errors are divided into two categories, duplications and definitional errors. Duplications, by the Census Bureau definition, occur when the same person is counted twice within a small geographic area, usually a set of neighboring blocks and enumeration districts. In most cases, the duplications occurred when the same housing unit was counted twice. For example, in rural areas there are no street addresses, and boundaries are often unclear. A house might be described in the Master

Address Register by the materials it was built from. When a follow-up enumerator saw the same house, perhaps unsure of its location on a map, (s)he would describe it by its color and shape, add the address to the MAR, and re-enumerate the household. Because the descriptions differ, no one would notice the same house was listed twice and two census forms would be collected. The Census Bureau (1985) reports that the rate of housing unit duplication was almost twice as high in rural as in urban areas.

Definitional errors fall into two main categories. One category consists of fabrications, in which the census-taker, perhaps despairing of ever finding anyone at home at a given address, simply made up the census information. The Bureau estimates that there were just under a million of these in 1980. The second category includes people who were counted at an address other than the one where they lived on April 1, 1980. Some of these were foreigners only visiting the United States. Others were people who moved between April 1 and the end of the census-taking period. Data collection lasted several months, through September or October in many areas, so a family moving after April 1 could be counted at the destination or perhaps at both origin and destination. Definitional errors also occurred when a person maintained two addresses, one perhaps a vacation home or an apartment in a distant city used by one member of a commuter marriage. It is also likely that many college students were counted both in their dormitories and by their parents at home. Many of the definitional errors are in fact duplications. They differ from the Census Bureau definition of a duplication because the two addresses where the person was counted are not located close together.

It should be noted that the Census Bureau estimate of nearly one million fabrications is an upper bound. The Bureau identified erroneous enumerations by means of a sample survey of 100,000 "counts". If a supposedly counted person could not be found by a survey interviewer and was not known to either neighbors or the local post office, (s)he was considered to be a fabrication. The Bureau felt that some of the Post Office checking may have been poor and that at least some of the people listed as fabrications were in fact properly counted. To the extent this is true, the correct estimate of erroneous enumerations is less than 6 million.

Most erroneous enumerations occurred during the follow-up phase of the census. In Table 1, I show rates of erroneous enumeration by census-taking method. We see that the rates of erroneous enumeration are much lower in conventional areas. Because no Master Address Register was compiled in advance, there was much less chance for housing unit duplication. In the mailout areas, we see that rates of erroneous enumeration are three to four times as great in the follow-up phase, where people were counted by enumerators, as in the mailback phase, where people counted themselves. Many of these duplications are the result of "coverage improvement" activities. Of the 2.6 million people added through these activities, an estimated 420,000 (U.S. Bureau of the Census, 1987) were duplicates.

Rates of erroneous enumeration varied across geographic areas. In Table 2, I present rates for central cities, remainders of SMSA, non-metropolitan urban areas (populations over 2,500), and non-metropolitan rural areas. The geographic patterns vary by type of erroneous enumeration. Definitional errors were more likely to occur in cities and duplications were more likely to occur in rural areas. Combining categories, rates of erroneous enumeration were highest in cities and rural areas and lower elsewhere. If we limit ourselves to those 16 cities with individual PEP estimates where the combined omissions rate was 9.7 percent, the rate of erroneous enumeration was 3.8 percent. In the 20 states where at least 10 percent of the population was Black or Hispanic, the rate was 3.7 percent and, in the remainder of the country, it was 2.7 percent.

In considering the non-metropolitan part of the country, it is important to separate conventional from mailout-mailback areas. This is especially true for rural areas, but I do not have access to the necessary data. Instead, I have combined non-metropolitan urban and rural areas to compare the effects of the two data collection methods (Table 3). Clearly, use of the mailout-mailback method in non-metropolitan areas increased the incidence of erroneous enumerations substantially. In 1970, almost all of the non-metropolitan areas in the United States were enumerated by the conventional method. The change in method artificially inflated the 1970-80 growth rate in such areas. This is especially true for Blacks where the rate of erroneous enumeration in rural areas was 7.1 percent, but only 3.3 percent elsewhere (Woltman, Alberti and Moriarity; 1988).

In Table 4, I compare the distributions of omissions and erroneous enumerations by type of geographic area, relying on tabulations by Fay (1988b). I present omission rate estimates for both April and August, using the two series, 3 and 5, with the same strategy for treating missing data. For erroneous enumerations, I exclude those geocoding errors which are not duplicates. We see in the table that the distributions of omission rates are similar, so the timing of data collection did not matter very much. Rates of erroneous enumeration are consistently lower, and do not show a large range of variation among areas. For omissions, rates are highest in large central cities and lowest in the largely suburban "other urban places" subcategory in the remainders of SMSA. The range of rates is 4.7 percent for April and 3.9 percent for August. For erroneous enumerations, the corresponding range is only 1.4 percent. I conclude that erroneous enumerations do not balance omissions by geographic area.

V. THE PATTERN OF NET UNDERCOUNT

As Table 4 suggests, to complete the study of geographic undercount differentials, we must consider erroneous enumerations and imputations as well as omissions. Some erroneous enumerations and imputations balance omissions. One person may be curbstoned (erroneous enumeration) into a unit where three actually live (omissions). One unit in a duplex may be counted twice while the other is missed entirely. A mover may be counted at the destination (erroneous enumeration) rather than the origin (omission). A passerby may say that four people (imputations) live in a house when the correct number is three (omissions). However, for erroneous enumerations and imputations to cancel out omissions, they must balance in every area. To examine the extent to which this has happened I present (Table 5) net undercount rates based on data from PEP Series 3-8. Net undercount rates separating all central cities and remainders of SMSA from non-metropolitan areas were unavailable, but I could define four types of area:

a. Sixteen large central cities;
b. Remainder of the SMSA's of these cities;
c. Seven additional SMSA's where separate results for central cities were not available; and,
d. The remainder of the country separated by region.

The results show a substantial net undercount in central cities, but rates in other areas are consistently low, rising above one percent only in the seven SMSA's (1.1 percent) and the West (1.9 percent), where the conventional method was used extensively (18 percent of the population).

Why does the pattern exist? It is partly due to the concentration of minorities in central cities. Nearly half (45 percent) of the combined population of the 16 cities is either Black or Hispanic, compared to 15 percent elsewhere. Blacks and Hispanics are also harder to count in the cities. For Blacks, the net undercount was 9.1 percent in the cities and 4.3 percent elsewhere. The corresponding rates for Hispanics were 7.9 and 3.4 percent. This shows that if a synthetic adjustment is likely to improve the census, we can do even better with a survey based adjustment recognizing higher minority undercount rates in cities.

VI. THE POSSIBILITY OF ADJUSTMENT

The results so far create a desire for adjustment. The question is whether available data will support it. To adjust the 1980 Census, we must rely upon PEP data which are admittedly imperfect and troubled by missing information.

For 4 percent of cases in the PEP omissions samples, Bureau statisticians could not decide whether a person was counted or omitted. For these cases, a count/omission status was imputed by computer, and a substantial proportion of the missing data cases were deemed to have been omitted. To understand the effect of this, we need to review the PEP procedure which included three steps. First, sample members of the April and August, 1980 editions of the Current Population Survey were matched against the census. Matching cases were assigned a final status of "counted". Other cases were not assigned a final status, and a field check was mandated. Bureau interviewers tried to visit each unresolved case to determine whether the person had in fact been counted. The interviews were not always successful, due to bad addresses, people moving away, or unwilling respondents. In the third step, final statuses for cases still unresolved were assigned by imputation. Donors for imputation were cases with similar demographic characteristics who had been unresolved at the end of the first step but were resolved in the second. Since many of the donors were omissions, 51 percent of the unresolved cases in the April sample were imputed to be omissions as well. The corresponding proportion for the August sample was 42 percent (Cowan and Bettin, 1982). Looking at it another way, 35 percent (April) or 28 percent (August) of the all omissions received this status by imputation.

The main alternative, known as "Series 14" was to discard the unresolved cases from the analysis and to adjust for nonresponse by increasing the weights assigned to resolved cases. This had the same effect as assigning the status of "omitted" to 5 percent of the unresolved cases and the status of "counted" to remaining 95 percent, and it reduced the omissions rate from 5.40 to 3.66 percent. To compare net undercount rates, we use the "Series 8" assumptions for erroneous enumerations. Substituting PEP Series 14-8 for 3-8, the net undercount rate for the nation fell from 0.8 to -1.0 percent, an overcount.

The Bureau in fact made five different assumptions for omissions and three different assumptions for erroneous enumerations. As shown by Ericksen, Kadane and Tukey (1987), the choice between omissions Series 3 and 14 is the one that really matters. Choices among assumptions for erroneous enumerations had little effect. Selecting a PEP series has been a subject of controversy, both at the 1984 New York Census trial and afterward (Freedman and Navidi, 1986). We believe that the assumptions for Series 3 are reasonable and take heart from the fact that the national undercount rates it provides are close to these obtained by demographic analysis. For the total population, demographic analysis provides an estimate of 1.4 percent, and for Blacks, 5.9 percent. For PEP series 3-8, the corresponding rates are 0.8 and 5.2 percent. The Black rate for Series 14-8 was 0.7 percent.

Few serious students of the undercount would opt for Series 14. Even so, many feel uneasy about the data, because so many omissions were assigned by imputation. Some observers are uncertain even about the matching decisions that were made in the field without imputation. Although interviews were taken, a certain amount of inexactness arises from the fact that the field check was made several months after both the Census and the CPS had taken place.

We could take the attitude that we expect measurement to be most difficult among the hard-to-count. Recent research (Boone, 1987; Fein and West, 1988) shows that people with weaker attachments to those they live with, e.g., distant relatives or boarders, are less likely to be counted. It stands to reason that such people might also be missed on the CPS or the field check, unresolved at the end of the second step, and properly imputed to be omitted. Moreover, some of the census "matches" occurring on the first step could be mistakes which would incorrectly deflate the estimated omissions rate.

A second, contrary, attitude points out that matching is especially difficult in poor neighborhoods of large cities. If it is more difficult to match than to count in such areas, then many people actually included in the census count would be mislabeled as "unresolved" after the second step. If such people lived in a suburb or rural area they would properly be matched to the census. This bias has the effect of overestimating the number of unresolved cases in cities, consequently inflating the final omissions rate estimate. Because this bias is less evident in suburbs or rural areas, use of the Series 3 imputation strategy would overestimate the undercount differentials.

To my knowledge no one has studied the differential accuracy of matching and imputation across areas. To decide about adjustment in 1980, we must limit ourselves to robustness tests. Do the various PEP series, which hopefully span the range of reasonable missing data strategies, lead to about the same result? If all PEP Series, regardless of the missing data assumptions, show higher undercount rates in cities, then some adjustment is called for. This is especially true for minority populations, since the White undercount rates are not especially variable across geographic areas. In Table 6, I compare the minority undercount rates in cities and elsewhere for all 12 PEP Series.

We see that regardless of the PEP series chosen, Blacks and Hispanics were harder to count in central cities than elsewhere. The range of difference varies, from 6.05 percent for Series 2-20 to 0.79 percent for Series 14-8, but all results lead to the same conclusion. Minority populations are more difficult to count in cities, and we would improve upon the results of a synthetic estimation by taking this into account. The remaining question is how much adjustment should there be? For this, we need to use subjective judgement in two ways, to select (1) a PEP series or combination of series; and (2) a model for estimation.

VII. MODELING THE ADJUSTMENT

Many statisticians have expressed concern about how to model the adjustment because there are too many choices. As Nathan Keyfitz put it in the New York Census trial:

> "[A]djustments could be devised, any one of which might make the figures better overall. And in that lies the principal difficulty: There is no way of choosing among the methods, and each would lead to different results for the 50 states and 39,000 local areas in competition for congressional seats and federal funds (1983, p. 5)."

Undercount research is a new field, and statisticians have not yet devised a method for selecting one best adjustment. As a problem, adjusting the census should be no more difficult than it is to estimate complicated quantities such as the Consumer Price Index, postcensal changes in per capita income and population, imputed values for missing data items on the census, and the number of undocumented aliens living in the United States. For such projects, some recourse to substantive knowledge and subjective judgement is needed. Adjusting the census is no different.

In the paragraphs below, I describe some of the modeling alternatives, indicating strengths and weaknesses. The data are sample estimates of undercount for 66 areas, 16 central cities, 12 remainders of states in which the cities are located and the 38 other whole states. One possibility is to use the sample estimates. I discard this for two reasons: (1) we would like to reduce the sampling error, and (2) we would like to know that our estimates are consistent with some theory of the undercount. For Series 3-8, the sample estimate for South Carolina was 6 percent and for Boston, it showed a 1 percent overcount. These seem implausible.

A second alternative is to use the combined sample estimates shown in Table 5, calculating one adjustment for central cities, and one each for the six other areas indicated. There are two problems with this approach: (1) counting conditions in central cities other than the 16 with separate PEP estimates might also be difficult, and (2) we want to give some weight to uniquely local conditions. The sample estimate for South Carolina might be out of line, but unique problems in the census offices located there could have caused the undercount rate to be somewhat higher than in neighboring states.

A third possibility is to use a simple empirical Bayes procedure. Here we take a group of areas where census-taking conditions are similar and calculate the mean sample estimate. If we assume that the sample estimates X_i estimate the corresponding population values θ_i, and that:

a. The θ_i are distributed normally with mean μ and variance σ^2, and

b. The X_i are distributed normally with mean θ_i and variance τ^2,

then we can estimate the sum of $\sigma^2 + \tau^2$ with $V^2 = \Sigma (X_i - \bar{X})^2/(p - 3)$, where p is the number of local areas. To estimate the undercount for area i, we calculate a weighted average of $\bar{X}$ and X_i. If we write s^2 as the sample estimate of σ^2, then the weight for $\bar{X}$ is s^2/V^2, and the weight for X_i is $(V^2 - s^2)/V^2$. We can see that where sample estimates are imprecise, and s^2 is large, more of the weight will be attached to the mean estimate. Where the sample estimates are precise, there will be little advantage to pooling in this manner.

The method can be illustrated for the 16 central cities, where a pooled estimate for s^2 is $(.02)^2 = .0004$. For the data shown in Table 5, $V^2 = .000599$, and the weights are .668 for the mean and .332 for the sample estimates. The composite estimates preserve the rank ordering, but all estimates are shrunk toward the mean, sometimes by substantial amounts as in Boston and San Diego.

This approach preserves all local idiosyncracies, and requires only one simple assumption, that the 16 areas have common census-taking conditions. The simplicity of the assumption creates its own difficulty. We need to argue that the 16 cities are the proper set. Why not separate the 11 cities where the proportion Black or Hispanic is more than 40 percent? Alternatively, why not include the 16 remainders of SMSA attached to these cities, since they are highly urban areas (See Freedman and Navidi, 1986)? Judgement is required to select the proper set, and the controversy surrounding the adjustment issue makes some statisticians hesitate to use their judgement. The choice of areas matters, as I show in Table 7. The estimates obtained for cities when we pool across 32 areas, the 16 central cities and 16 remainders of SMSA, are consistently lower with the average reduced to 2.61 from 3.50.

Thinking about factors causing the undercount discourages the use of the simple empirical bayes procedure. It is appealing to think of the undercount as systematically related to common causes, such as the percent minority, the confluence of difficult living conditions represented by a high crime rate, or the manner in which the census is taken. These hypotheses led Ericksen and Kadane (1985; 1987) to suggest using a hierarchical Bayes procedure in which a regression estimate is substituted for the mean estimate just discussed. The resulting composite estimate averages the regression and original sample estimates. Use of these estimates requires a longer list of assumptions (see Freedman and Navidi, 1986 for a critical review), but they also permit us to see whether the undercount rates are systematically related to conditions thought to reflect difficult counting conditions. We draw comfort from the fact that among large cities, those with more Blacks and Hispanics have higher sample estimates, and that crime rates, proportion of persons with language difficulties, and the concentration of poverty are all systematically related to the undercount rates. Ericksen, Kadane and Tukey (1987) obtained a good regression fit with just three independent variables, and the regression estimates got much larger weights than the sample estimate when the composites were calculated.

Referring to Table 7, we see that the hierarchical Bayes procedure, here calculated over all 66 local areas, preserves the overall mean for the 16 cities (3.57 compared to 3.50). Even so, the estimates sometimes diverge greatly from both the sample estimates and the empirical Bayes estimates. The three predictor variables in the regression equation are the crime rate and the percentages minority and conventional. Cities like Boston and Saint Louis, where crime rates are especially high, have high composites, although the sample estimates are low or moderate. In contrast, Chicago and Philadelphia have low crime rates, and the hierarchical Bayes estimates are much lower than either the sample estimates or empirical Bayes estimates.

The two models diverge because they depend on different information. In the empirical bayes case, I grouped areas which seemed to have similar census-taking conditions, namely large cities with many minorities. In the second case, I estimated an equation in which the dominant predictor variable was the crime rate, highly variable among the 16 cities, and related to the undercount rates of the 50 states outside the 16 cities. A better strategy would be to stratify the sample by the same variables you would use in the regression model. This was not possible in 1980 given the design of the PEP samples and available predictor variables. In 1990, there is a good chance that the problem will be corrected, because the PES sample is being stratified by variables shown to be related to undercount rates (Woltman, Alberti and Moriarity, 1988).

VIII. SUMMING UP

The divergence between empirical and hierarchical Bayes results will cause some statisticians to hesitate to choose between them. To reduce the possibility for controversy in 1990, I have three suggestions to make, some of which have already been adopted by the Census Bureau.

The first is to increase the sample sizes for local areas. The Bureau has already proposed a national sample size of 300,000 for the 1990 Post Enumeration Survey, nearly three times the size of the 1980 PEP samples. This increased sample size will have at least three beneficial results: (1) the local estimates will be more precise, reducing shrinkage toward a common mean or regression estimate and giving greater weight to idiosyncracies in local counting conditions; (2) separate estimates can be created for a larger number of areas, e.g., moderately large central cities, and (3) there will be fewer anomalous results like those observed in 1980 in Boston and South Carolina.

The second is to create more homogeneous sampling units. Most states are heterogenous areas with a wide variety of counting conditions. New Jersey, for example, includes the difficult-to-count minority neighborhoods in Camden, Jersey City, and Newark, large numbers of affluent suburbs, and many small towns and rural areas. By creating more homogeneous sampling units, which do not necessarily lie within state lines, we can better identify those factors which affect the undercount. This proposal has also been adopted by the Census Bureau (Woltman, Alberti and Moriarity, 1988).

The third, discussed above, is to stratify the sampling units by variables related to counting conditions. Examples of stratifying variables are location in cities, suburbs, or rural areas, whether the conventional or mailout-mailback method is used, the expected mailback rate, whether the census office is centralized or decentralized, and the predominant type of housing. The stratifying variables can be used either to define subsets of areas for empirical Bayes estimation or to serve as independent variables in regression.

To make an intelligent decision on adjustment we need to take all available information into account, giving greater weight to the things we think we know best. Having reviewed the Census Bureau reports on the subject, I believe we are quite certain of the following:

a. The net undercount in 1980 was substantial. If we assume that there were between 2 and 4 million undocumented aliens, the rate was between 0.9 and 1.8 percent.
b. The undercount was differential, with Blacks and Hispanics being missed at rates much higher than Whites.
c. Omissions and erroneous enumerations have different geographic distributions.
d. Use of a synthetic estimator based on age, race, and Black/non-Black status improves upon the unadjusted census.
e. Blacks and Hispanics had higher undercount rates in large cities than they did elsewhere.

We are less certain of the following propositions:

f. PEP Series 3 can be used to compute local estimates of net undercount.
g. We can select a model with which to adjust the census.

If we can agree to statements f and g, then we would opt to adjust the census. We would like to have more information on errors in the PEP data and a good theory of the undercount to base the selection of a model upon. To date, no one has made a study of bias in the use of imputation on the PEP. The robustness of the imputation rules could be tested by varying the matching variables and restricting donors to those cases resolved with greatest certainty.

The more interesting question for statisticians concerns the selection of a model. The model should be consistent with a theory of the undercount. The problem is that few statisticians, inside or outside the government, have really tried to develop a theory. Much of the relevant research has only been recently done and is not widely known. This theory, though, should be used to define the sampling for the Post Enumeration Survey and the grouping or independent variables for calculating estimates. When this is done, it is likely that the results of an empirical bayes analysis shrinking estimates toward a group mean would better agree with the results of a hierarchical bayes model shrinking toward a regression estimate.

Boone, Margaret (1987), "Inner City Black Undercount," *Evaluation Review*, 11, pp. 216 - 241.

Cowan, Charles D., and Bettin, Paul J. (1982), "Estimates and Missing Data Problems in the Post Enumeration Program," technical report, U.S. Bureau of the Census, October.

Cowan, Charles D., and Fay, Robert E. (1984), "Estimates of Undercount in the 1980 Census," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 566 - 571.

Ericksen, Eugene P., and Kadane, Joseph (1985), "Estimating the Population in a Census Year (with comments and rejoinder)," *Journal of the American Statistical Association*, 80, pp. 98 - 131.

Ericksen, Eugene P., and Kadane, Joseph (1987), "Sensitivity Analysis of Local Estimates of Undercount in the 1980 U.S. Census," in R. Platek, J.N.K. Rao, C.E. Sarndal, and M.P. Singh, *Small Area Statistics: An International Symposium*, New York: John Wiley, pp. 23 - 45.

Ericksen, Eugene P., Kadane, Joseph B., and Tukey, John W. (1987), "Adjusting the 1980 Census of Housing and Population," *Technical Report #401*, Department of Statistics, Carnegie-Mellon University.

Fay, Robert E. (1985), "Implications of the 1980 PEP for Future Census Coverage Evaluation," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 413 - 418.

Fay, Robert E. (1988a), "1980 Census Preliminary Results Memorandum No. 119," unpublished Census Bureau report, March 10.

Fay, Robert E. (1988b), "1980 Census Preliminary Results Memorandum No. 121," unpublished Census Bureau report, March 10.

Fein, David J., and West, Kirsten K. (1988), "Towards a Theory of Census Undercount: An Exploratory Assessment of Data from the 1986 Los Angeles Test Census," paper presented at U.S. Bureau of the Census, March.

Freedman, David A., and Navidi, W.C. (1986), "Regression Models for Adjusting the 1980 Census (with comments and rejoinder)," *Statistical Science*, 1, pp. 1 - 39.

General Accounting Office (1980), "An Assessment of 1980 Census Results in 10 Urban Areas," report number 66D-81-29.

Keyfitz, Nathan (1983), "Affidavit" presented to U.S. District Court, Southern District of New York, 80 Civ. 4550.

National Research Council (1985), *The Bicentennial Census: New Directions for Methodology in 1990.* Washington, D.C.: National Academy Press.

Schirm, Allen L., and Preston, Samuel (1987), "Census Undercount Adjustment and the Quality of Geographic Population Distributions (with comments and rejoinder)," *Journal of the American Statistical Association*, 82, pp. 965 - 990.

Schoen, Douglas (1980), Affidavit presented to U.S. District Court, Southern District of New York, 80 Civ. 4550.

United States Bureau of the Census (1975), "Coverage of Population in the 1970 Census and Some Implications for Public Programs," *Current Population Reports*, Series P-23, No. 56, Washington, D.C.: U.S. Government Printing Office.

United States Bureau of the Census (1982), "PEP Estimates for States, and Sums of State Estimates to Regional and National Totals: PEP Estimates for Substate Areas," submitted to U.S. District Court, Southern District of New York, 80 Civ. 4550.

United States Bureau of the Census (1985), "The Coverage of Housing in the 1980 Census," *Evaluation and Research Reports. 1980 Census of Population and Housing*, PHC80-E1, Washington, D.C.: U.S. Government Printing Office.

United States Bureau of the Census (1987), "Programs to Improve Coverage in the 1980 Census," *Evaluation and Research Reports. 1980 Census of Population and Housing*, PHC80-E3, Washington, D.C.: U.S. Government Printing Office.

United States Bureau of the Census (1988), "The Coverage of Population in the 1980 Census," *Evaluation and Research Reports. 1980 Census of Population and Housing*, PHC80-E4, Washington, D.C.: U.S. Government Printing Office.

Woltman, H. Alberti, N., and Moriarity, C. (1988), "Sample Design for the 1990 Post Enumeration Survey," paper presented at the Annual Meeting of the American Statistical Association, Section on Survey Research Methods.

TABLE 1

RATES OF ERRONEOUS ENUMERATIONS BY TYPE BY CENSUS-TAKING METHODOLOGY

Type of Erroneous Enumeration	Mailed Back Questionnaires	Enumerator Collected Questionnaires	Mailout Area Total	Conventional
	(Percent)			
	(1)	(2)	(3)	(4)
Duplications	0.79%	2.87%	1.17%	0.21%
Definitional Errors	1.06	4.36	1.65	1.12
Post Office Cases[1]	0.20	1.61	0.45	0.24
Other[2]	0.86	2.76	1.20	0.88
Total	1.85	7.24	2.82	1.34
Total Including Geocoding Errors	2.47	8.22	3.50	1.65

Sources and Notes

[1] Consists of cases where no one, including Post Office contacts, knew of the person included in the census. It is thought that the majority of these cases are fabrications.

[2] Consists of cases where people were counted at addresses other than their usual place of residence on April 1, 1980.

Source: Cowan and Fay (1984).

TABLE 2

RATES OF ERRONEOUS ENUMERATION BY TYPE OF AREA

Type of Erroneous Enumeration	Metropolitan Areas: Central Cities	Metropolitan Areas: Balance	Non-Metropolitan Areas: Urban	Non-Metropolitan Areas: Rural
	(Percent)			
	(1)	(2)	(3)	(4)
Duplications	1.02%	1.08%	0.94%	1.74%
Definitional Errors	2.06	1.36	1.62	1.48
Post Office Cases[1]	0.78	0.36	0.28	0.13
Other[2]	1.28	1.01	1.34	1.35
Total	3.07	2.45	2.56	3.22
Total Including Geocoding Errors	3.38	3.17	3.20	4.48

Sources and Notes

1 Consists of cases where no one, including Post Office contacts, knew of the person included in the census. It is thought that the majority of these cases are fabrications.

2 Consists of cases where people were counted at addresses other than their usual place of residence on April 1, 1980.

Source: Fay (1988a).

TABLE 3

RATES OF ERRONEOUS ENUMERATION IN NON-METROPOLITAN AREAS

Type of Erroneous Enumeration	Conventional	Mailout-Mailback	Ratio (2)/(1)
	(Percent)		
	(1)	(2)	(3)
Duplications	0.21%	1.61%	7.7
Definitional Errors	1.12	1.66	1.5
Total	1.34	3.27	2.4
Total Including Geocoding Errors	1.65	4.42	2.7

Sources and Notes

Sources: Cowan and Fay (1984); Fay (1988a).

TABLE 4

DISTRIBUTIONS OF ERRONEOUS ENUMERATIONS AND OMISSIONS BY TYPE OF AREA

Type of Area	Omission Rates Series 3[1]	Omission Rates Series 5[2]	Rates of Erroneous Enumeration Series 8[3]
	(Percent)	(Percent)	(Percent)
	(1)	(2)	(3)
Central Cities			
Population of 500,000 or more	8.8%	8.7%	3.6%
Population Less Than 500,000	6.1	6.7	2.7
Remainders of SMSA			
Urban Places, Population of 25,000 or more	4.3	5.1	2.2
Other Urban Places	4.1	4.8	2.3
Rural Balance	4.9	5.7	3.1
Non-Metropolitan Areas			
Urban	4.7	5.4	2.6
Rural	5.3	6.7	3.2

Sources and Notes

1 Based on April, 1980 edition of the Current Population Survey.
2 Based on the August, 1980 edition of the Current Population Survey.
3 Excluding geocoding errors which are not duplicates.

Source: Fay (1988b).

TABLE 5

RATES OF NET UNDERCOUNT BY GEOGRAPHIC AREA

Area	Rate (Percent)
16 Central Cities[1]	4.5%
16 Remainders of SMSA[2]	0.9
7 Whole SMSA's[3]	1.1
Remainders[4]	
Northeast	-1.1
North Central	0.8
South	0.2
West	1.9

Sources and Notes

1 The cities are Baltimore, Boston, Chicago, Cleveland, Dallas, Detroit, Houston, Indianapolis, Los Angeles, Milwaukee, New York City, Philadelphia, Saint Louis, San Diego, San Francisco and Washington, D.C.
2 These are the remainders of SMSA that are left when the cities listed in note 1 are removed.
3 The SMSA's are Atlanta, Denver, Kansas City, Miami, Newark, Pittsburgh and Seattle.
4 Remainders refer to the areas of four census regions that are left when the SMSA's listed in notes 1, 2 and 3 are removed.

Source: United States Bureau of the Census (1982), PEP Series 3-8.

TABLE 6

MINORITY[1] NET UNDERCOUNT RATES FOR CENTRAL CITIES AND ELSEWHERE, 12 PEP SERIES

PEP Series	Central Cities[2]	Elsewhere	Difference
	(Percent)		
	(1)	(2)	(1)-(2) (3)
2 - 8	9.14%	4.27%	4.87%
2 - 9	10.98	5.14	5.84
2 - 20	11.56	5.51	6.05
3 - 8	8.76	3.95	4.81
3 - 9	10.61	4.82	5.79
3 - 20	11.19	5.19	6.00
5 - 8	6.68	4.75	1.93
5 - 9	8.58	5.60	2.98
10 - 8	4.80	2.54	2.26
14 - 8	1.29	0.50	0.79
14 - 9	3.29	1.38	1.91
14 - 20	3.91	1.76	2.15

Sources and Notes

1 Minorities include Blacks and non-Black Hispanics.
2 The cities are Baltimore, Boston, Chicago, Cleveland, Dallas, Detroit, Houston, Indianapolis, Los Angeles, Milwaukee, New York City, Philadelphia, Saint Louis, San Diego, San Francisco and Washington, D.C.

Source: United States Bureau of the Census (1982).

TABLE 7

ESTIMATES OF NET UNDERCOUNT FOR 16 CITIES BY FOUR DIFFERENT MODELS

	Model			
City	Sample Estimates[1]	Empirical Bayes Cities[2]	Empirical Bayes Cities and Suburbs[3]	Hierarchical Bayes[4]
	(Percent)			
	(1)	(2)	(3)	(4)
Baltimore	5.35%	4.11%	3.39%	4.46%
Boston	-1.00	2.01	0.74	4.58
Chicago	4.36	3.79	2.97	2.92
Cleveland	4.91	3.97	3.20	4.05
Dallas	5.93	4.31	3.63	4.77
Detroit	3.07	3.36	2.43	4.87
Houston	4.60	3.87	3.07	3.02
Indianapolis	-0.18	2.28	1.08	0.58
Los Angeles	4.56	3.85	3.06	3.95
Milwaukee	3.14	3.38	2.46	1.74
New York City	6.04	4.34	3.67	4.35
Philadelphia	4.75	3.92	3.14	2.17
Saint Louis	3.13	3.38	2.46	5.66
San Diego	-0.96	2.02	0.75	1.78
San Francisco	4.64	3.88	3.09	3.32
Washington, D.C.	3.61	3.54	2.66	4.96
Mean of 16 Estimates	3.50	3.50	2.61	3.57

Sources and Notes

1 Obtained from PEP Series 3-8.
2 Obtained by empirical bayes procedure over 16 central cities.
3 Obtained by empirical bayes procedure of 32 areas - 16 central cities and the 16 remainders of SMSA. The grand mean for all 32 areas was 1.98.
4 Obtained from a weighted average of sample estimate and regression equation with three independent variables - the crime rate, the percent Black or Hispanic, and the percent of population counted by the conventional method.