Should We Have Adjusted the Census of 1980?

By

D A Freedman
Statistics Department
University of California
Berkeley, CA 94720

and

W C Navidi
Mathematics Department
University of Southern California
Los Angeles, CA 90089

Technical Report No. 306
June 1991

Department of Statistics
University of California
Berkeley, California 94720

Should we have adjusted the census of 1980?

by

D A Freedman
Statistics Department
University of California
Berkeley, Ca 94720
Research partially supported by NSF Grant DMS 86-01634

and

W C Navidi
Mathematics Department
University of Southern California
Los Angeles, Ca 90089

Abstract

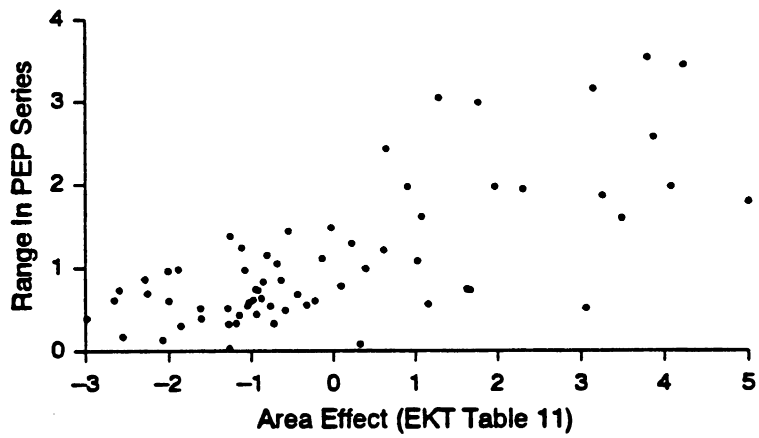
This paper reviews some of the arguments for and against adjusting the census of 1990, and the decision of the court.

Keywords and phrases

Census, adjustment, Post Enumeration Survey, regression, smoothing

To appear in Survey Methodology

1980 Census Paper, Figure 1, June 10/91, pepneg.ps, dot=1.00
Freedman, 415-642-2781



1. Introduction

Every ten years, the census gives a statistical portrait of the United States. Geographical detail makes these data unique. However, the counts have more than academic interest: they influence the distribution of power and money. The census is used to apportion Congress as well as local legislatures and to allocate tax money--$40 billion per year in the late 1980s--to 39,000 state and local governments. For these purposes, the geographical distribution of the population matters, rather than counts for the nation as a whole. Indeed, the census is used as a basis for sharing out fixed resources: if one jurisdiction gets more, another must receive less. Adjusting the census is advisable only if the process brings us closer to a true picture of the distribution of the population.

A small undercount is thought to remain in the census, and this undercount is unlikely to be uniform. People who move at census time are hard to count; in rural areas, maps and address lists are incomplete. Central cities have heavy concentrations of poor and minority persons, who may be harder to enumerate. If the undercount can be estimated with good accuracy, especially at the local level, adjustments can--and should-- be made to improve the census. Some statisticians argue that the undercount can be estimated well enough, others are skeptical: a bad adjustment may be worse than nothing.

Because of its resource implications, the undercount has attracted considerable attention in the media, the Congress, and the courts. After the 1980 census, New York City joined with other jurisdictions to sue the Department of Commerce, seeking to compel an adjustment based on demographic analysis and capture-recapture techniques. The Commerce Department resisted this pressure. The trial court framed the issue as follows:

"The plaintiffs contend that a statistical adjustment of the census will improve upon the accuracy of the census, thereby reducing the disproportionate undercount in the City and State [of New York]. The Census Bureau, however, contends that although the census counts are imperfect, a statistical adjustment of the census will inject even greater inaccuracies into the population count, and that therefore, a statistical adjustment of the census is not technically feasible or warranted at this time." [674 *F Supp* 1091= volume 674 of the Federal Supplement, page 1091].

The 1980 case may seem dated, given that the census of 1990 has already been taken. However, among law suits that involve statistical principles, the 1980 census case was one of the most important and closely argued; there is still much to learn from it. This article will review some of the technical issues, and some of the findings of the court.

The balance of this section will sketch the background; for more details, see Cohen and Citro (1985) or Fay et al. (1988). There are two methods for evaluating the completeness of the counts in the U.S. Census: demographic analysis and capture-recapture. Demographic analysis uses administrative records (birth certificates, death certificates, immigration visas, etc.) to make independent estimates of population totals. The starting point is an accounting identity:

Population = Births - Deaths + Immigration - Emigration.

Demographic analysis provides estimates by age, sex and race but not ethnicity, because of gaps in the records. Data on immigration and emigration are incomplete; birth records are incomplete too, especially prior to 1935. Thus, the data going into the "identity" must be supplemented by a variety of imputations and adjustments. Furthermore, data on internal migration are lacking, so estimates are made primarily at the national level. This completes our sketch of demographic analysis.

Estimates of coverage for small areas (including states and cities) are based on capture-recapture techniques. Capture is in the census; recapture is in a sample survey conducted after the census. In 1980, there were two such surveys, or "P-samples:" the April and August CPS (Current Population Survey). Each record from the P-samples was matched against the census file to see if the corresponding person was "captured," that is, counted in the census. Records that could not be matched indicated people who were missed by the census-- or a failure in the matching process. These data were used to estimate the percentage of persons missed by the census, that is, the rate of omissions.

The census also had a small percentage of erroneous enumerations (for instance, people counted at two different addresses); the number was estimated by taking an "E-sample" of census records and trying to check them by field work. In effect, the net undercount was estimated by taking the difference between the omissions and erroneous enumerations. (For details, see Fay et al. Chapter 5.) These undercount estimates were made as part of "PEP," the Post Enumeration Program.

In 1980, there was a fair amount of missing data in the P- and E-samples: for instance, there was a 4% non-interview rate in the CPS; even after interview, a determination of match status could not be made for another 4% of the subjects. To see the effect of missing data, a variety of imputation schemes were considered, leading to 12 different series of PEP estimates for 66 subareas.

The 66 areas covered the whole U.S. They included cities like New York; states apart from these cities, like upstate New York; and whole states like Wyoming. A PEP "series" consists of 66 estimates, one for each study area; 9 of the 12 series were based on the April CPS, and 3 on the August CPS.

In the 1980 case, expert witnesses for plaintiffs included Gene Ericksen, Jay Kadane, and John Tukey. Their strategy for adjusting the census using PEP data was described in Ericksen and Kadane (1985). Freedman (among other statisticians and demographers) testified for the defendants, and Navidi was a consultant. A critique of the proposed adjustments was summarized in Freedman and Navidi (1986), to be referenced here as F&N.

We now indicate some of the technical issues. According to experts from the Bureau of the Census:

a) There were substantial differences among the 12 PEP series, demonstrating that missing data were a serious problem.

b) The PEP estimates were subject to large biases, apart from the problems created by missing data.

c) Each PEP series was subject to unacceptably large sampling error.

Ericksen and Kadane responded that one of the PEP series ("PEP 2-9") was preferred, and that sampling error could be substantially reduced by regression modeling. They proposed a model with two equations. The first equation expresses the idea that y_i, the PEP estimate for study area i, is an unbiased estimate of the true undercount γ_i for that study area. Informally,

PEP estimate for area i = True undercount in area i + Random error

Formally,

(1) $y_i = \gamma_i + \delta_i$.

The second equation expresses a theory about the variation of the undercounts from area to area, in terms of a vector of explanatory variables X_i and a vector of hyper-parameters β. Informally,

True undercount in area i	=	Linear combination of explanatory variables for area i	+	Random error

Formally,

(2) $\gamma_i = X_i \cdot \beta + \varepsilon_i.$

The assumptions on the error terms can be stated as follows:

(3) $E(\delta_i) = E(\varepsilon_i) = 0.$

(4) $\text{var } \delta_i = K_i, \text{ var } \varepsilon_i = \sigma^2.$

(5) $\delta_1, \delta_2, \ldots, \delta_{66}, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{66}$ are independent.

(6) δ_i and ε_i are normally distributed.

In (4), K_i is the split-sample variance for y_i computed by the Bureau; randomness in K_i is ignored; σ^2 does not depend on i and is treated as constant even though it is estimated from the data. The role of assumptions, and departures from them, was examined in F&N; also see the discussion papers and rejoinder, as well as sections 6-7 below.

The Ericksen-Kadane model was used in the 1980 case to smooth the PEP estimates, with the objective of reducing sampling error. The main focus of F&N was a critique of that model. Ericksen, Kadane and Tukey (1989)-- to be referenced here as EKT-- replied to F&N, and the present paper continues the exchange.

EKT cited a paper by Schirm and Preston (1987), which considers adjusting states and the District of Columbia by the "synthetic method." For instance, demographic analysis (with one set of assumptions on illegal immigration) estimated a national undercount rate of 5.9% for blacks and 0.7% for whites in 1980. The synthetic method adjusts each state as follows: increase the number of blacks by 5.9% and the number of whites by 0.7%. In short, undercount rates are assumed to depend on race but not geographical area-- or anything else.

This completes our summary of the technical background. For an update on the 1990 census, see Freedman (1991); some of the introductory material here was excerpted with minor changes from that paper. For other views, see Hogan and Wolter (1988), Schirm (1991), Wolter (1991), Wolter and Causey (1991), or Ericksen, Estrada, Tukey and Wolter (1991). The balance of the present paper responds to the salient points raised by EKT, and indicates how some of the the conflicting views were resolved by the trial court.

2. Do the adjustments improve on the census?

The most important question is whether adjustments improve on the census counts. EKT "...are confident of improving upon the raw census count [p943]"; indeed, there are

"two simple [synthetic] adjustments that improve upon the census....the question of the Ericksen and Kadane model is not whether it proves that adjustment is feasible, but whether it improves upon the simpler methods [pp927-8].... Study of the method will not 'prove' that an adjustment will improve the census. This has already been demonstrated by Schirm and Preston and the results of Tables 5 and 6 [p933]."

Thus, EKT's Tables 5 and 6 are the main pieces of empirical evidence to show that adjustment will improve on the census. And Table 6 on erroneous enumerations is redundant, because the PEP estimates in Table 5 include the effect of erroneous enumerations. Table 5 is the critical one, and it is reproduced here for ease of reference. In our opinion, the table says very little about the possibility of improving on the census; to see why, some numerical detail is needed. (Schirm and Preston will be discussed in the next section.)

Table 1. EKT's Table 5. Changes in National Population Shares Resulting When Counts Are Adjusted by Sample Estimates Pooled Across Areas And Synthetic Estimates. [The entries for the three groups represent changes in shares, or differential undercounts; the entries in the last column represent total undercounts.]

PEP estimate	Group 1	Group 2	Group 3	Estimated national undercount rate
2-20	+.52%	+.09%	-.61%	+1.9%
3-20	+.51%	+.08%	-.59%	+1.7%
2-9	+.50%	+.06%	-.56%	+1.6%
3-9	+.49%	+.04%	-.53%	+1.4%
2-8	+.41%	+.04%	-.45%	+1.1%
3-8	+.39%	+.03%	-.42%	+1.0%
5-9	+.31%	+.25%	-.56%	+2.1%
5-8	+.22%	+.23%	-.45%	+1.7%
14-20	+.21%	+.02%	-.23%	-.2%
10-8	+.19%	+.07%	-.26%	+.3%
14-9	+.19%	-.01%	-.18%	-.5%
14-8	+.10%	-.03%	-.07%	-1.0%
Synthetic A	+.17%	+.14%	-.31%	+1.4%
Synthetic B	+.12%	+.06%	-.18%	+1.4%
Shares of Census count	10.76%	44.24%	45.00%	

Notes. (i) Group 1 includes 16 central cities. Group 2 includes three state remainders (California, Maryland, and Texas, excluding Group 1 cities) and 17 whole states. All areas are at least 10% Black or Hispanic. Group 3 includes nine state remainders and 21 whole states. All Group 3 areas are less than 10% Black or Hispanic.

(ii) The Synthetic A estimates assume that (a) Blacks have the same undercount rates as Hispanics, 5.9%; (b) the undercount rate of persons neither Black nor Hispanic is 0.3%; (c) the undercount rates for Blacks, Hispanics, and all others are invariant across geographic areas; and (d) there are 3 million undocumented aliens, 9.6% of whom are Black.

(iii) Following Schirm and Preston (1987), the Synthetic B estimates assume that (a) the Black undercount rate is 5.9%; (b) Hispanics and other non-Blacks have an undercount rate of .7%; (c) the undercount rates for Blacks, Hispanics, and all others are invariant across geographic areas; and (d) there are 3 million undocumented aliens, 9.6% of whom are Black.

"Group 1" in the table consists of 16 central cities; "group 2" consists of other study areas that have relatively high minority populations; "group 3" consists of study areas with small minority populations. At best, the table shows that several methods for adjusting these groups are in general agreement. The table does not show that any of the methods improve on the accuracy of the census. It cannot, because there is no external standard against which to measure improvement.

Moreover, we believe the impression of agreement in the table to be largely illusory. There are dramatic differences among EKT's preferred PEP series, or between these series and the synthetic adjustment of Schirm and Preston. Of course, drama depends on scale, and our next task is choosing units. Proponents of adjustment often use "loss functions" to make their argument; squared error is a common choice: see Ericksen, Estrada, Tukey and Wolter (1991, p20). EKT view Schirm and Preston as demonstrating census adjustment to be advantageous, so we compute the root mean square difference between the census and the "Synthetic B" line in Table 1, which is based on the Schirm and Preston adjustment. (The mean is weighted by population shares.)

$$\sqrt{.11\times(.12)^2 + .44\times(.06)^2 + .45\times(.18)^2} \approx 0.13 \text{ of } 1\%.$$

In short,

(7) rms difference between census and synthetic B = 0.13 of 1%

EKT prefer the first 8 of the PEP series (pp933 and 938). We next compute the rms difference between PEP 2-20 and 3-8, which are among EKT's preferred series. (PEP 2-20 and 3-8 were both based on the April CPS; differences between them are due only to procedures for handling missing data.)

(8) rms difference between PEP 2-20 and 3-8 = 0.14 of 1%

EKT also recommend averaging as a way of eliminating indeterminacies (pp931 and 937). Table 2 compares population shares from the census, the synthetic B estimates, and the average preferred PEP estimates. We take the rms difference between the average preferred PEP and synthetic B:

(9) rms difference between
average preferred PEP and synthetic B = 0.25 of 1%

Table 2. Population shares from the census, the synthetic B estimates, and the average of EKT's eight preferred PEP series (2-20, 3-20, 2-9, 3-9, 2-8, 3-8, 5-9, 5-8).

	Group 1	Group 2	Group 3	Total
Ave. Preferred PEP - Synthetic B	.30%	.04%	-.34%	.00%
Census - Synthetic B	-.12%	-.06%	+.18%	.00%
Average Preferred PEP	11.18%	44.34%	44.48%	100.00%
Synthetic B	10.88%	44.30%	44.82%	100.00%
Census	10.76%	44.24%	45.00%	100.00%

A comparison of (7-8-9) reveals three salient points:

a) the difference between the census and synthetic B is rather small;

b) the range in the preferred PEP series is larger than the difference between the census and synthetic B;

c) the difference between the average preferred PEP and synthetic B is twice the difference between the census and synthetic B.

EKT must view a difference of 0.13% as serious: see (7). On this scale, the PEP series do not agree among themselves. Furthermore, the PEP series are very different from the synthetic adjustment. Of course, the reason may be that Schirm and Preston did not go far enough. However, a National Academy of Sciences review panel-- with Jay Kadane as a prominent member-- reached the tentative conclusion that Schirm and Preston already over-adjusted the census: see Cohen and Citro (1985, p.287).

The PEP estimates are in better agreement with the "synthetic A" adjustment in Table 1. But this is circular: the undercount rate for hispanics in synthetic A was estimated from PEP, while synthetic B was based on demographic analysis. Differences among the PEP estimates are an awkward reality; and so are differences between the PEP estimates and synthetic adjustments.

We now quote the principal claim made by EKT (p927):

"Our conclusion is that regardless of whether we use one of the simple methods or the composite method and regardless of how we vary the assumptions of the composite method, an adjustment reliably reduces population shares in states with few minorities and increases the shares of large cities."

Giving more money to cities by changing the census counts is a good idea only if the adjustment reliably improves the accuracy of the census. Accuracy is the crucial issue, and we wish EKT would address it more directly. Their Table 5 is almost irrelevant.

3. Schirm and Preston

Can synthetic adjustment reliably improve on the accuracy of the census? EKT think so, citing Schirm and Preston (1987) for the evidence. Schirm and Preston present two major arguments, one analytic and one based on simulation. However, both have serious flaws.

The analytic argument (p966):

"Our finding is that synthetic adjustment will always move the estimated ratio of a state's population to the national population closer to the true ratio if

a) the state's black undercount is closer to the national black undercount than it is to the national undercount for both races combined and

b) the state's white undercount is closer to the national white undercount than it is to the national undercount for both races combined."

As a matter of mathematics, this proposition is wrong. A counter-example is given in Table 3: state A, for instance, has by construction 89 whites and a census count of 90.

Table 3. A counter-example to the analytical argument. There are two states and two races.

	White		Black		Total	
	Census count	True count	Census count	True count	Census count	True count
State A	90	89	1	2	91	91
State B	910	890	99	119	1009	1009
Total	1000	979	100	121	1100	1100

The counter-example has been set up to make the arithmetic easy; more complicated and realistic examples could undoubtedly be provided. In Table 3, the overall error in the census (white plus black) is 0, for each state and for the nation. Thus, the census gets the state shares right, and any adjustment will make matters worse. Error rates (with the true population as base) are shown in Table 4: Schirm and Preston's conditions are satisfied. Synthetic adjustment moves both states farther from truth, as shown in Table 5; state B is helped, state A is hurt. To compute Table 5 from Table 3, the number of whites in state A is multiplied by

(10) true national total for whites/national census total = 979/1000.

The arithmetic for the other cells is similar.

Table 4. Undercounts from Table 3, in percent. (Negative undercounts correspond to overcounts.)

	White	Black	Total
State A	-1.1%	50%	0%
State B	-2.2%	17%	0%
Total	-2.1%	17%	0%

The counter-example may be informative, as a parable: state A is sparsely populated, with a small minority population; state B is heavily populated, and has a large, hard-to-count minority population. Synthetic adjustment may favor states of type B at the expense of type A. The mathematical error in Schirm and Preston's appendix appears to be in their reasoning from display A.2. Professor Preston informs us (personal communication) that the theorem holds, with a more complicated set of conditions involving weighted averages.

Table 5. The synthetic adjustment, "Syn."

	White		Black		Total	
	Syn	True count	Syn	True count	Syn	True count
State A	88	89	1	2	89	91
State B	891	890	120	119	1011	1009
Total	979	979	121	121	1100	1100

This completes our discussion of the analytic reasoning in Schirm and Preston. What about the simulation results? Basically, Schirm and Preston consider 51 areas (the states and D.C.) and two races (black and white). They set up a joint distribution for an assumed "true" population and the census counts; both are taken as stochastic. The census counts can be adjusted by the synthetic method, and the question is whether the raw counts or the adjusted counts are closer to the assumed true counts. Schirm and Preston actually consider several joint distributions, defined by different "scenarios," that is, choices of parameters; the results are quite similar across scenarios. They also consider several loss functions, or measures of closeness.

We focus on Scenario I, and make two brief comments.

a) The claimed improvement is rather modest. For example, on average, just over half the population lives in states whose shares are made more accurate by adjustment-- no matter how small the improvement.

b) The "true" population was constructed on the basis of the synthetic assumption-- no systematic variation in undercount rates within race across geography; random variation was allowed. See equation (2) in Schirm and Preston. Thus, the definition of "truth" favors synthetic adjustment.

On the whole, however, Schirm and Preston have a reasonable argument. If the assumptions of the synthetic method more or less hold, its estimates will be good. There remains the crucial question: do those assumptions hold? what kind of geographical variation is there in undercount rates? On this score, Schirm and Preston offer no evidence. In the 1980 case, the trial court found that

"the synthetic method simply ignores geographical variations and assumes that a person is as likely to be missed in the census whether he lives in Alabama or in Alaska. However, as defendants' experts persuasively explained, this assumption that the undercount rates for the various age, race, and sex groups are constant from one subnational area to another has no basis in fact whatsoever.... the synthetic method is simply inadequate as a means of adjusting the census."
[674 *F Supp* 1098, footnotes and citations omitted]

4. Adjusting small areas

Statistical adjustment of census counts is more likely to be beneficial at fairly high levels of geographical aggregation (for instance, census regions or divisions). However, there are 39,000 state and local governments in the U.S., all claimants for tax money. Many of these jurisdictions are further subdivided, into city council seats, etc. If census counts are to be adjusted, they must for legal and policy reasons be adjusted at quite fine levels of geographical detail. Indeed, the proposal for 1990 is to adjust down to the block level. (A "block" is the smallest unit of census geography; there are 6.5 million blocks in the U.S.)

EKT discuss two synthetic methods for adjusting subareas of the 66 study areas, as well as a regression method (p941). In the end, however, there is no evidence that adjustment of small areas will improve on the raw census counts. With respect to 1980, EKT say (p943):

"For the 66 areas included in our study, we are confident of improving upon the raw census count, especially in those areas with large undercounts or overcounts where an adjustment is most needed. Our findings do not permit definitive conclusions for suburban areas, for central cities other than the 16 included in our data set, or for other rural or urban parts of individual states. To compute estimates for such areas, we would prefer not to extrapolate from the regression equations presented in this article."

EKT go on to describe alternative designs for capture-recapture sampling, leaving open the question of small-area adjustment for 1990. Much of the dispute in 1980 centered on the feasibility of adjusting small sub-areas of the 66 study areas. To win its case, New York had to show such adjustments would improve on the census. EKT now seem to concede there was little evidence on this score.

5. Averaging and sensitivity analysis

The 12 PEP series were the results of a sensitivity analysis on missing data. Since the amount of missing data was large relative to the undercount, methods for handling missing data have impact. In response, EKT offer quite a variety of procedures for adjusting the census on the basis of the various PEP series, including: a) eliminating discrepant series (pp937-9); b) eliminating systematic differences between the series (pp937-8); c) regression on other variables (the "composite" estimator, pp933ff); d) averaging (pp931 and 937).

This list makes clear the essential indeterminacy of census adjustment schemes. And in this context, the use of averages to reduce indeterminacy needs discussion. Arbitrary modeling decisions may be defensible if they do not matter-- the usual robustness argument. Sensitivity analysis (changing the assumptions to see if the results change) may refute the robustness argument. However, averaging the results from a sensitivity analysis is self-defeating. The different PEP series are not repeated measurements of the undercount. It is the spread in the PEP series that is interesting, not the average-- because it is the spread (among, say, the April series) that demonstrates the impact of different modeling assumptions on the same data.

6. Assumptions

EKT (p937) say the model improves on the PEP estimates and the synthetic method. The model does improve on the PEP estimates, if you grant its assumptions-- equations (1-6) above. So far, however, these equations still seem quite implausible. Likewise, the model improves on the synthetic estimates only if it uses the additional variables in a sensible way, bringing us right back to assumptions.

At times, EKT seem to argue that the model can be inferred from the data (pp933ff). Of course, there is more to a regression model than choice of variables on the left hand side and the right hand side, although that is difficult enough, as will be seen below. There are many questions to answer: Why are effects linear and additive-- equations (1-2) above? What about the assumptions on the errors-- equations (3-4-5-6)? And so forth. EKT put forward no evidence to justify their assumptions, except by attempting to rebut our rebuttal (p931). Do they think a model is right unless it can be proved wrong?

In any case, we stand by our critique. For some data on correlation bias, see Fay et al. (1988, esp. sec. 6F); for a critique of Ericksen and Kadane's estimates, see Fellegi (1985, p.118). Other sources of bias in the PEP series include matching errors and errors in census-day address reports.

EKT argue that PEP is "conservative" (p931). This seems to be both wrong and irrelevant; wrong because the biases generally increase the apparent undercount: and irrelevant because geographical variation in the biases matters a great deal. Assumption (3) is rather unlikely: the errors probably do not have mean 0. The undercounts estimated by PEP are likely to be biased upward, the size of the bias depending on the area. For a review of the evidence, see Fay et al., chap. 6; also see F&N. The trial court in the 1980 case concluded:

"The evidence at trial established that the PEP was plagued by various errors caused by inadequacies in the PEP methodology. This type of error is referred to as 'bias.' A significant source of bias in the PEP arises because the process of matching people from the CPS to the census...is an extraordinarily difficult and inexact task. Because of inaccurate, irregular, and incomplete information in both the CPS and the census, the Bureau undoubtedly and inevitably made many errors in determining the match status of individuals enumerated in the CPS, thereby distorting the P sample's undercount estimate. Moreover, the evidence at trial established that most of this matching error occurred because the Bureau erroneously determined many cases to be misses when they were in fact matches. This error, therefore, resulted in the PEP overstating the undercount. The extent of this error and the degree to which it varies from one geographic area to another is unknown." [674 *F Supp* 1100, footnotes and citations omitted]

We turn now to equations (4-5). Take the independence assumption. In 1980, there were 3 processing offices and 12 regional offices. EKT's counter: there were 400 district offices. Granted. There were also several dozen area managers, several hundred thousand census staff and about 1500 CPS interviewers. The sources of error are numerous, and dependence seems likely. Processing offices, regional offices, managers, census interviewers, and CPS interviewers all must contribute components of error, to say nothing of respondents. Likewise, the constancy of σ^2 in (4) seems unlikely: different parts of the country are undercounted for different reasons, not readily captured in a linear regression equation.

We pointed out that random events like snowstorms might cause correlated errors in several areas; EKT respond that there were no snowstorms. This issue goes to the foundations of statistics: if the weather is good, the errors are independent; but in foul weather, all bets are off. The distributions in the model, and the statistical inferences, are therefore conditional on certain events. Which ones, and why?

Fortunately, we do not need to resolve the problem of conditional vs. unconditional inference. There was a major event that disrupted census operations over several states in the Pacific Northwest. Mt. St. Helens erupted in May 1980, while followup interviewing was in full swing.

7. Other issues

7.1 Does it matter which series is used?

At the level of precision EKT demand of the census, the different PEP series-- even among their preferred ones-- really do lead to quite different adjustments, as shown by equations (7-8-9). EKT, however, claim that the preferred PEP series all lead to similar adjustments. And to support their position they offer Table 11, which suggests for example that New York City has a differential undercount of 3.27% with an uncertainty of 0.62%.

For many purposes, a uniform undercount would not be material; it is differential undercounts that create inequities. The "area effects" seem to be measures of differential undercount-- the policy variable of main interest.

The "area effects" in the table were computed by EKT as follows:

i) Restrict attention to 8 of the 12 PEP series.
ii) Smooth each of these using the regression model.
iii) For each area, take the average of the 8 estimates.
iv) Subtract the corresponding national estimate of undercount.

Table 6 below compares "area effects" with differences in the PEP estimates, attention being restricted to the preferred series based on the April CPS. Differences among these PEP estimates are due only to differences in the handling of missing data. Taking the range seems fair: reasons for data to be missing can differ from area to area, and so will the appropriate imputation procedure. Adding in the August series would increase the range, but some of the difference would be due to sampling error.

The table shows that for some areas, the effects are large relative to differences between PEP series, suggesting that missing data have little impact on the results. Upstate New York is an example. But for other areas, like Chicago, the reverse holds and imputation procedures matter.

Table 6. Comparing area effects with differences in the PEP estimates, restricted to preferred series based on the April CPS. Subareas match those used in F&N.

	Preferred April PEP series			Area effect
	Min.	Max.	Range	
Alabama	-.37	.60	.97	-1.07
Alaska	2.79	3.53	.74	1.63
Los Angeles	4.56	7.72	3.16	3.16
San Diego	-.98	1.45	2.43	.65
San Francisco	4.31	6.25	1.94	2.31
Rest of California	2.84	3.92	1.08	1.03
Chicago	3.57	6.56	2.99	1.77
Rest of Illinois	1.21	1.75	.54	-1.04
New York City	6.04	7.90	1.86	3.27
Rest of New York	-1.61	-1.44	.17	-2.55
Wyoming	3.91	4.04	.13	1.16

All 66 areas are plotted in Figure 1. The x-axis shows the area effect; the y-axis shows the range in the preferred April PEP series. In root mean square (across the 66 areas), the spread among EKT's preferred PEP series-- based on the April CPS-- is about 75% of the area effect. In other words, the impact of missing data (never mind other biases in PEP) is similar in magnitude to the effect EKT are trying to measure. Bringing in alternative imputation models would make matters even worse. Nor is averaging the results a good fix, for reasons given earlier.

The positive association in Figure 1 is quite striking, and so is the change in the joint distribution when the area effect changes from negative to positive. Our explanation: PEP estimates of undercount are indicators of poor data quality-- in PEP as well as the census. Large apparent undercounts indicate areas with poor data. In such areas, there is a lot of missing data, so the effect of changing the imputation rules will be large too. Areas that are hard to count are also hard to adjust. See F&N p9 or Wolter (1986, p26, points 8 and 9).

There may be some reasonable way of choosing a compromise version among the PEP series. But why are any of the PEP series, or their averages, an improvement over the census? That is the crucial question, and EKT do not answer it. In our view, adjustment--whether by a synthetic method, or a PEP series, or a regression model, or any convex combination--will in the end be driven mainly by assumptions.

Figure 1. PEP and data quality. For each of the 66 study areas, the horizontal axis shows the EKT "area effect." The vertical axis shows the range in the preferred April PEP series.

7.2 Which PEP series is best, and which explanatory variables should be used?

At trial, and in their discussion of F&N, Ericksen and Kadane recommended an adjustment based on PEP 2-9, apparently the most preferred of all 12 series. We chose PEP 10-8 as an alternative for study. EKT defend 2-9, and try to exclude 4 of the series-- especially our foil 10-8. The arguments were reviewed in court and in F&N (p8, the discussion, and the rejoinder p36). Our opinion remains the same: there is no rational basis for choosing 2-9 over 10-8.

EKT impute to us the position that "proportion urban" should have been considered as an independent variable (p934). This is not quite right. We felt that EKT's choice of independent variables was somewhat arbitrary, and wanted to show that changing variables made a real difference to the results-- another sensitivity analysis. The difference was observed mainly for small areas (F&N, p9). Since EKT no longer advocate adjustment of small areas in 1980, this argument may be moot.

There is one new twist to the reasoning: EKT argue for choosing models by "reliance on statistical criteria (p941)." In essence, they recommend choosing variables so as to minimize the rms residual in an OLS fit. However, the rms residual measures association in the data not correctness of underlying theory.

For reasons that remain unclear, EKT restrict attention to models with 2, 3, or 4 variables; and they require coefficients to have t-statistics of 2 or more. Their preferred equation seems to be

$$\text{(11)} \quad \text{PEP 2-9} = \underset{(-4.0)}{-2.23} + \underset{(5.4)}{.079}\ \text{min} + \underset{(3.6)}{.036}\ \text{crime} + \underset{(3.5)}{.028}\ \text{conv} + \text{residual}$$

rms residual = 1.53

The right hand side variables are the percent minority in the study area, the crime rate, and the percent conventionally enumerated; t-statistics are shown in parentheses; the rms residual is computed using the unbiased divisor n-p. This equation is used only for variable selection; after the variables are chosen, the model is refitted by GLS: see (1-6) above, and F&N for discussion.

The statistical logic is not apparent, and EKT's criteria have to be read quite literally. For example, here is another candidate equation:

$$\text{(12)} \quad \text{PEP 2-9} = \underset{(7.6)}{.120}\ \text{min} + \underset{(3.4)}{.026}\ \text{crime} + \underset{(3.8)}{.029}\ \text{conv} - \underset{(-4.4)}{.176}\ \text{pov} + \text{residual}$$

rms residual = 1.49

The additional variable is the percentage of persons in the study area with incomes below the poverty level; the intercept was suppressed because the t-statistic was small. Equation (12) fits a little better than (11) in terms of rms residual, and "shows" that the undercount goes down as the percentage of poor people goes up--other things being equal. EKT reject this equation because the coefficient of *pov* is significantly negative rather than significantly positive.

Preconceptions about the undercount may be incompatible with the data, and best-subsets OLS may not be a suitable analytic technique. We reject neither interpretation, but our main conclusion is this. In the present context there are no objective, statistically defensible criteria for model selection. Much rides on the subjective judgment of the modeler.

With this in mind, we return to the points at issue--choosing a PEP series, and deciding between the crime rate or the percent urban as explanatory variables. As far as we can see, on the criteria chosen by EKT, the difference between crime rate and percent urban is trivial. And PEP 10-8 is clearly better than 2-9. See Table 7.

Table 7. Rms residuals from regression equations for PEP 2-9 and PEP 10-8. Explanatory variables include percent minority, percent conventionally enumerated, and either the crime rate or the percent urban.

	Crime rate	Percent Urban
PEP 2-9	1.53	1.54
PEP 10-8	1.35	1.33

On pp935 and 940 of EKT, σ denotes the rms residual. There is some conflict in notation, because we wrote σ^2 for Var(ε) in equations (2) and (4), following Ericksen and Kadane (1985, p105) or F&N (p5). To avoid conflict, let SE(ε) be the estimated value for our σ; this is what controls the standard errors of the 66 area undercounts computed by the Ericksen-Kadane model, as shown by equations (8) and (10) in F&N. For PEP 10-8, the estimated SE(ε) is virtually 0, so a model based on 10-8 fits extremely well and the 66 area undercounts are very precisely estimated (Table 8).

Table 8. SE(ε) and the rms SE for the 66 study areas; PEP 2-9 and PEP 10-8. The models include percent minority, percent conventionally enumerated, and either the crime rate or the percent urban.

	Crime rate		Percent Urban	
	SE(ε)	rms area SE	SE(ε)	rms area SE
PEP 2-9	.75	.65	.76	.65
PEP 10-8	.00	.28	.00	.25

Notes. Let K be a 66×66 diagonal matrix, whose (i,i) element is K_i. Let X be the 66×4 matrix of explanatory variables. Let $H=X(X^TX)^{-1}X^T$ and $\Gamma^{-1} = K^{-1} + SE(\varepsilon)^{-2}(I-H)$. The 66 area undercounts are estimated by the Ericksen-Kadane model as $\Gamma K^{-1}y$, where y is the 66×1 vector of PEP estimates. The rms SE for the 66 study areas is $\sqrt{\text{trace } \Gamma/66}$. For details, see F&N. At trial, Ericksen and Kadane estimated SE(ε) from 51 study areas (whole states and DC); we followed suit in F&N. Here, we use the 66 study areas, since that seems to be EKT's current recommendation. The difference is noticeable.

On "statistical criteria," contrary to the claims made by EKT, 10-8 is preferred to 2-9 and percent urban is just as good an explanatory variable as the crime rate. Their qualitative critique seems off the mark too. Of course, different urban areas are different, just as EKT say. So are different central cities. Similarly, minority persons living in central cities are likely to be different from those in suburbs. And so forth. All of EKT's variables are "blurred predictors" of undercount, and some are blurrier than the percent urban (p934).

With respect to this set of issues, the judge in the 1980 case was harder on Ericksen and Kadane than we are:

"Moreover, as defendants' experts persuasively explained, no one series of PEP estimates can be reliably shown to be superior to the others, or indeed, to the census itself, because there is insufficient knowledge with respect to which PEP procedures are better suited for measuring census undercount. While two of plaintiffs' experts expressed a preference for the 'series 2-9' PEP estimates based upon the hypothesis that the PEP procedures employed in arriving at those estimates were superior to the procedures used for the other PEP estimates, the plaintiffs' experts offered nothing more than unsupported assumptions in support of that position. On the other hand, the defendants' experts offered equally plausible assumptions which favored different PEP procedures, producing dramatically different PEP estimates." [674 *F Supp* 1102, footnotes and citations omitted.]

7.4 Simulation studies

We had a simulation study making three points: a) you could not infer from the data which variables go into the model, b) standard errors depend on assumptions about disturbance terms, and c) the standard errors computed by Ericksen and Kadane were quite optimistic. We had two additional points on this topic: d) standard errors do not measure the impact of bias; e) the Ericksen-Kadane smoothing simply passes through any bias in PEP that is well related to the explanatory variables.

Points a) through e) are real obstacles to showing that the model improves on the PEP estimates. EKT do not comment on points b), d) and e). They deny a), but more or less concede point c). For our part, we concede that in our simulation-- which grants half the model-- regression does reduce sampling error. We still think a) is right, as will be argued below. And in other contexts, smoothing may actually increase sampling error (Ylvisaker 1991 p7).

EKT (p943) criticize our study, because it covered only models with three variables in the equation and did not restrict the t-statistics. So we repeat the simulation here. In essence, we take PEP 10-8 as "truth," and add for each of the 66 study areas i a random error with variance K_i, as in (4). This grants equation (1) and the assumptions on δ_i. We choose variables according to the procedure outlined by EKT (p935), and fit the regression model, repeating the whole process 100 times.

Table 9 shows the variables selected in the first 10 runs. As will be seen, there is no consistency-- except that the percentage "conventionally enumerated" always comes in. Over the 100 runs-- excluding the ones that produced no acceptable model-- the nominal rms error was about 30% too small, and improvement of the composite estimator over PEP was exaggerated by a factor of 1.75. Assumptions matter.

A minor digression on census procedures. "Conventional enumeration" means that respondents were asked to fill out the forms and hold them for collection by an enumerator; this process was used in largely rural areas, particularly in the west. *Conv* is the percentage of persons living in areas that were conventionally enumerated. (In urban areas, forms were to be mailed back.) The undercount in 1980 was relatively high in rural areas, probably due to incomplete maps and address lists; that may be why *conv* is such a powerful explanatory variable.

We did an additional simulation with PEP 2-9 taken as truth, allowing percent urban to be selected as an explanatory variable. The results are shown in Table 10. Again, the percent conventionally enumerated comes in as does the percent minority. Otherwise, there is a fair degree of inconsistency. And the much-maligned percent urban is chosen more often than 5 of EKT's variables, including the central-city indicator. The data do not determine the model.

Table 9. A simulation experiment on variable selection; PEP 10-8 is taken as "truth."

Run	CC	Min	Crime	Conv	Ed	Pov	Lang	MU
1			x	x	x			
2		x		x				
3		x		x				
4	x		x	x				
5	x			x				
6		x		x				
7			x	x	x			
8		x		x				
9	x			x				
10	There was no model satisfying EKT's criteria							

Notes. *CC* is an indicator for central cities; *Min*, the percentage of minorities; *Crime*, the crime rate; *Conv*, the percentage who were conventionally enumerated; *Ed*, the percentage with no high school degree; *Pov*, the percentage below the poverty line; *Lang*, the percentage who have difficulty with English; *MU*, the percentage living in multiple-unit housing.

Table 10. A simulation experiment on variable selection. PEP 2-9 is taken as "truth"; percent urban (*Urb*) is permitted as an explanatory variable. The table shows the number of times each variable is entered, and the average of its coefficient (over the times it enters); 100 data sets were generated.

Variable	No. of times entered	Average coefficient
CC	17	2.954
Min	82	0.071
Crime	53	0.053
Conv	93	0.028
Ed	5	0.085
Pov	1	0.135
Lang	17	0.315
MU	0	*****
Urb	23	0.060

7.5 The regression model at trial

As statisticians, we are intrigued by arguments about regression. However, the court was not impressed:

"In their rebuttal case, the plaintiffs argued that the application of regression analysis to the undercount estimates derived from the PEP would enable the Bureau to use the PEP to accurately adjust the 1980 census. However, both plaintiffs' and defendants' experts agreed that regression analysis will not in any way alleviate the bias in the PEP and plaintiffs apparently do not contend otherwise. In short, while regression analysis may remove some of the random sampling error in the PEP, regression analysis will not reduce the substantial errors in the PEP caused by erroneous matches, the untested assumptions made with respect to the unresolved cases, and correlation bias. Moreover, the overwhelming weight of the evidence supports the conclusions of defendants' experts that the principal difficulties with the PEP stem from these biases rather than from sampling error." [674 *F Supp* 1103, footnotes and citations omitted.}

8. Summary and conclusion

Ericksen, Kadane, and Tukey argue that they can improve on the 1980 census counts by statistical adjustment. They seem now to agree that adjustments would not have been justified for subareas of the 66 PEP study areas. With respect to the 66 areas themselves, disagreement remains. In our opinion, success of any of EKT's proposed adjustments rides on unverified and implausible assumptions-- about missing data, undercount mechanisms, bias in PEP, and stochastic errors in regression models. Changing the assumptions changes the results, and taking averages over various sets of assumptions does not, at least in our opinion, make the problem go away.

EKT conclude (p943),

"We believe that the Census Bureau creates political difficulties for itself when it ignores the undercount. The bureau will put itself in a better position by making its best effort, using available statistical and demographic methods, to adjust for the undercount. Errors will remain, but they will be smaller and we will no longer know in advance who is losing money and power because of the undercount."

This political analysis has merit, but there are caveats. We think it quite unfair to say that the Bureau has ignored the undercount. Nor are the Bureau's political difficulties entirely of its own creation. Adjustments can indeed be devised to satisfy particular groups or settle individual law suits. However, the census is used to share out fixed resources, so there will always be losers as well as winners. These will have little trouble identifying themselves, after the fact if not before. And up to now, the goal of improving on the accuracy of the census by statistical adjustment has proved illusory.

9. How did the court rule?

At the time of writing, litigation about the 1990 census goes on. With respect to the 1980 census, however, the court ruled for the defendants on all the issues. We quote from the digest and opinion *Cuomo et al. v. Baldrige et al.* 674 *F. Supp.* 1089-1108 (SDNY, 1987).

"State, city, and their officials brought action against Secretary of Commerce, Director of the Bureau of the Census, and other officials seeking statistical adjustment of 1980 decennial census. The District Court, Sprizzo, J., held that state and city failed to establish that statistical adjustment of decennial census was technically feasible."

"....it is essential to any such adjustment that a technically feasible adjustment methodology exist which gives a truer picture of the United States population on a state-by-state basis for apportionment purposes, and a sub-state-by-sub-state basis for federal funding purposes....If it does not, then no adjustment can or should be made...because...both congressional seats and revenue sharing funds are fixed quantities, and an increase in the population in one state or sub-state area will adversely affect the shares of other localities....

"Notwithstanding the complexity of the facts....this action presents one issue to be resolved by the Court: whether the plaintiffs have sustained their burden of proving that a statistical adjustment of the 1980 census will result in a more accurate picture of the proportional distribution of the population of the United States on state-by-state and sub-state-by-sub-state basis than the unadjusted census. The Court finds as a matter of fact that the plaintiffs have not sustained that burden, and the action must therefore be dismissed...."

Acknowledgments. Freedman is working for the Department of Justice on matters arising from the 1990 census. However, the views expressed in this article may not be shared by the Department. Helpful comments were made by L. Bazel (San Francisco), P. Diaconis (Harvard), S. Klein (RAND) and A. Tversky (Stanford).

Appendix. Synthetic estimation and loss functions

Synthetic estimation

Section 5 in Wolter & Causey (1991) describes their empirical proof that synthetic adjustment would have brought the 1980 census closer to truth. The evidence is a simulation study: the "census" and "truth" are both defined in terms of an artificial reference population developed by Isaki et al (1987). However, the argument depends rather strongly on the reference population, as shown by Passel (1987). The object here is to sketch a variation on one of Passel's examples. Indeed, if the reference population is defined by using PEP 2-9 to correct the 1980 census, then synthetic adjustment moves the counts farther from truth.

Table 11 shows the data for the four census regions--Northeast, Midwest, South, and West. With squared differences in population shares weighted by size,

(13) r.m.s. difference between Synthetic B and PEP 2-9 = 0.21 of 1%

(14) r.m.s. difference between the Census and PEP 2-9 = 0.15 of 1%

Table 11. Population shares from the census, the Synthetic B estimates, and PEP 2-9, in percent; census counts, in 1,000s.

	Northeast	Midwest	South	West	Total
Synthetic B - PEP 2-9	.08%	.03%	.24%	-.35%	.00%
Census - PEP 2-9	.10%	.06%	.12%	-.28%	.00%
PEP 2-9	21.59%	25.92%	33.15%	19.34%	100.00%
Synthetic B	21.67%	25.95%	33.39%	18.99%	100.00%
Census	21.69%	25.98%	33.27%	19.06%	100.00%
Census count	49,135	58,866	75,372	43,172	226,545

PEP 2-9 is rather close to the "average preferred PEP" in Table 2. In that table, the census was closer to synthetic B than to the PEP estimates. In Table 11, the census is closer to PEP, and synthetic B is the outlier. The difference between the two tables seems to be the disaggregation. Table 2 disaggregates the U.S. by race and ethnicity; Table 11, according to conventional census geography.

Of course, using another disaggregation or a different synthetic adjustment could reverse the comparisons yet again; so could a change in the loss function. To illustrate the possibilities, consider adjusting the 66 PEP study areas, rather than four regions. Keep PEP 2-9 as 'truth.' Using the loss function (17), the census is preferred to synthetic B, by a little. Using (16), synthetic B shows a much smaller loss than the census.

Loss functions

Proponents of adjusting the 1990 census make analytic arguments based on loss functions: see Wolter & Causey (1991) or Ericksen, Estrada, Tukey & Wolter (1991, p20 of the main report; Appendices G & H). The essence of argument can summarized in the lemma which follows. To set up the notation: the country is divided into n areas, indexed by i; c_i is the census count in area i and t_i is the true count. The "synthetic estimate" for area i is $x_i=\lambda c_i$, where the "adjustment factor" λ is computed from other data.

(15) Lemma. For $i=1,\ldots,n$, let $c_i>0$ and $t_i>0$. Let $0<\lambda<\infty$ and $x_i=\lambda c_i$. Then

(16) $$\sum_{i=1}^{n}(x_i-t_i)^2/c_i$$

is minimized when $\lambda = \left[\sum_{i=1}^{n} t_i\right] / \left[\sum_{i=1}^{n} c_i\right]$.

The proof is omitted as trivial. The "loss function" defined by (16) differs in detail from the one used in (7-8-9) and (13-14), which can be written as

(17) $$\sum_{i=1}^{n} \frac{c_i}{C}\left[\frac{x_i}{X} - \frac{t_i}{T}\right]^2, \text{ with}$$

$$C = \sum_{i=1}^{n} c_i, \quad X = \sum_{i=1}^{n} x_i, \quad T = \sum_{i=1}^{n} t_i.$$

The loss function (17) emphasizes shares while (16) emphasizes counts; furthermore, (17) puts more weight on large sub-populations while (16) does the opposite, due to the division by c_i. We are not particularly attached to (17), and see no good way to choose one loss function rather than another.

Lemma (15) is mathematically correct, but it is so far removed from the realities of adjusting the 1990 census that it seems virtually irrelevant. In this connection, there are four points to consider:

a) The true population total T is unknown; Wolter & Causey attempt to deal with this problem, but the example in Table 11 refutes their argument: synthetic adjustment makes the 1980 census less accurate.
b) Synthetic estimates do not perform well under aggregation.
c) At the block level, rounding error may dominate.
d) Loss functions only capture part of the policy problem, and may obscure more than they reveal.

Points b-c-d) will be discussed in more detail; but first, a brief review of proposed methods for adjusting the 1990 census. The population is divided into 1392 "post strata," e.g., male hispanic renters age 30-44 in central cities in the Pacific Division. Index these post strata by $j=1,\ldots,1392$. For each post stratum j, an adjustment factor λ_j is computed by capture-recapture techniques from data collected in a Post Enumeration Survey (Freedman 1991).

The 1392 factors are used to adjust all small-area counts as follows. Fix an area, e.g., a town. This area will intersect many of the post strata. The census count for each area × post stratum intersection is multiplied by the corresponding λ_j, and the products are summed. In other words, subpopulations are adjusted by the synthetic method, and synthetic estimates are aggregated to obtain totals for small areas.

This completes a sketch of the adjustment process, and we return to points b-c-d).

b) Synthetic estimates do not perform well under aggregation. This was already pointed out by Fellegi. See Cohen & Citro (1985, p318). For another example, see Tables 3-5 above.

c) At the block level, rounding error may dominate. Census adjustment would in fact be done at the block level. (A "block" is the smallest unit of census geography; there are 6.5 million blocks in the country.) A typical block in an urban area may intersect 25 post strata; each block × post stratum intersection contains only a handful of people. Multiplying by an adjustment factor means adding or subtracting a fractional number of people, and the fractions would be rounded. The next example illustrates how rounding error may offset any advantage from synthetic adjustment.

Suppose there are n "areas" to adjust; these could be viewed as blocks intersected with one fixed post stratum. Suppose each of these areas has the same census count, c. Fix m<n. Suppose that in each of m areas, the census has missed one person; in the remaining n-m areas, the census count is exactly right. In all, there is an undercount of m people. These facts are considered as known; but it is not known which blocks have the missing people. According to (16),

(18) loss from using the unadjusted census = m/c

Adjustment would proceed as follows: choose m areas at random, and add one person to each of these areas. Clearly,

(19) expected loss from adjusting

$$= \frac{m}{n} \cdot m \cdot 0 + (1-\frac{m}{n}) \cdot m \cdot \frac{1}{c} + \frac{m}{n} \cdot (n-m) \cdot \frac{1}{c} + (1-\frac{m}{n}) \cdot (n-m) \cdot 0$$

$$= 2(1-\frac{m}{n}) \cdot \frac{m}{c}.$$

(18) Lemma. If m<n/2, there is an expected net loss from synthetic adjustment.

Proof. If m<n/2, then

$$2(1-\frac{m}{n}) \cdot \frac{m}{c} > \frac{m}{c}. \qquad \square$$

Of course, this example is almost as stylized as Lemma (13). In short, the value of census adjustment cannot be established by *a priori* argument.

d) Loss functions capture only part of the policy problem, and may obscure more than they reveal. To begin with an example, suppose that the census is in error, and the main impact of that error is to transfer a congressional seat from California to Pennsylvania. There is a gain for Pennsylvania, and a loss for California. There may be a net social loss from this misallocation, but attempting to quantify that loss by (16)-- or any similar formula-- seems quite simplistic.

We now present another example to illustrate point d). To focus the issue, suppose the census undercount is largely confined to blacks and hispanics in New York, Chicago, Houston, and Los Angeles. The census, by assumption, under-estimates the share of the population living in these four cities, and adjustment will partly correct that error.

Due to its reliance on the synthetic method, however, adjustment will change population shares everywhere. Areas which are heavily black and hispanic will have their population shares artificially increased, at the expense of other areas. This will be so even in regions of the country where the census was accurate.

In this example, the distribution of resources between the four cities and other areas may be made fairer by adjustment--at the expense of distortions introduced everywhere else. The loss-function approach slides over this difficulty. Balancing inequities is a political problem, not easily resolved by a statistical formula.

Some observers may consider the example to be extreme. However, the Post Enumeration Survey only samples 5000 blocks, and there are 39,000 jurisdictions to adjust. Real information about the undercount is necessarily confined to relatively few localities. Adjustments for other areas must therefore be based largely on theory rather than data.

References

C. F. Citro, M. L. Cohen, eds. (1985). *The Bicentennial Census: New Directions for Methodology in 1990.* (National Academy Press, Washington, D.C.)

E. P. Ericksen and J. B. Kadane (1985). Estimating the population in a census year. *Journal of the American Statistical Association* 80 98-131 (with discussion).

E. P. Ericksen, J. B. Kadane, J. W. Tukey (1989). Adjusting the 1980 census of population and housing. *Journal of the American Statistical Association* 84 927-43.

E. P. Ericksen, L. F. Estrada, J. W. Tukey, K. M. Wolter (1991). Report on the 1990 Decennial Census and the Post Enumeration Survey, submitted to the Secretary of the Department of Commerce, June 22, 1991.

R. E. Fay, J. S. Passel, J. G. Robinson, C. D. Cowan (1988). *The Coverage of the Population in the 1980 Census.* (U.S. Department of Commerce, Government Printing Office, Washington, D.C.).

I. Fellegi (1985). Comment. *Journal of the American Statistical Association* 80 116-9.

D. A. Freedman (1991). Adjusting the 1990 census. *Science* 252 1233-36. Copyright 1991 by the AAAS. Excerpted by permission.

D. A. Freedman and W. C. Navidi (1986). Regression models for adjusting the 1980 census. *Statistical Science* 1 1-39 (with discussion).

H. Hogan, K. Wolter (1988). Measuring accuracy in a post-enumeration survey. *Survey Methodology* 14 99-116.

C. Isaki, G. Diffendahl & L. Schultz (1987). Report on statistical synthetic estimation for small areas. Technical report, Bureau of the Census.

J. Passel (1987). A note about synthetic estimates of undercount. Memorandum, Bureau of the Census.

A. L. Schirm and J. Preston (1987). Census undercount adjustment and the quality of geographic population distributions. *Journal of the American Statistical Association* 82 965-90 (with discussion).

A. L. Schirm (1991). The effects of census undercount adjustment on congressional apportionment. *Journal of the American Statistical Association* 86 526-41

K. Wolter (1986). Comment. *Statistical Science.* 1 24-8.

K. Wolter (1991). Accounting for America's uncounted and miscounted. *Science.* 253 12-15.

K. Wolter and B. Causey (1991). Evaluation of procedures for improving population estimates for small areas. *Journal of the American Statistical Association* 86 278-84.

D. Ylvisaker (1991). A look back at TARO. Technical report, Department of Mathematics, UCLA.