Comment

Eugene P. Ericksen

1. INTRODUCTION

In 1980, I testified in behalf of New York City and State in their lawsuit against the Census Bureau. In that trial, I presented a theory of the undercount which argued that blacks and Hispanics were harder to count than whites, especially in central cities where census-taking problems accumulate (Ericksen, 1980). The effect of this undercount was made worse by overcounts elsewhere, especially in rural areas where inexact addresses made duplications more likely.

New York won this lawsuit, but the judgment was later reversed for procedural reasons by an appeals court. A new trial was ordered, which took place in early 1984. By this time, the Census Bureau had developed data which permitted the testing of my theory of the undercount. Demographic results were consistent with prior expectations. The black undercount was 4.8%, and the nonblack rate was −1.1%, an overcount (U. S. Bureau of the Census, 1982). Adding reasonable estimates of the numbers of undocumented aliens to the demographically estimated national total increased both rates a point or two, and the differential was narrowed slightly.

The Post Enumeration Program (PEP) was designed to supplement demographic analysis by providing survey-based undercount estimates for blacks, nonblack Hispanics, and others, for the nation, states, large cities, and metropolitan areas. The PEP showed a pattern consistent with demographic results. Blacks and Hispanics were indeed harder to count than whites, especially in those central cities with concentrations of minorities. The same pattern was found for all 12 series of PEP estimates produced by the Census Bureau. This is illustrated here by Series 2/8 (Table 1). There were 11 areas, all central cities, where the percentage of black or Hispanic was over 40%, and for these the black and Hispanic undercount rates were more than double the corresponding rates observed elsewhere.

There was little disagreement at the second trial, and there appears to be little disagreement today, that New York City and similar places were differentially undercounted. The question is whether adjustments to the census counts improve the situation. Jay Kadane and I argue that they can, and we suggest that

Eugene P. Ericksen is a member of the Department of Sociology at Temple University, Philadelphia, PA 19122.

TABLE 1
Estimated 1980 undercount rates by ethnicity and type of area

Black or Hispanic in area	Group			
	Blacks	Hispanics	Others	Total
%		%		
40 or more	10.1	8.5	1.0	5.3
10–39.9	4.9	3.8	0.2	1.2
0–9.9	3.4	3.0	−0.1	0.1
Total	6.1	4.7	0.1	1.1

Note: Areas are 16 central cities, 12 remainders of states in which the central cities are located, and 38 other states. There are 11 areas, each a central city, where the percentage of black or Hispanic is 40% or more. The source of the sample estimates is PEP series 2/8 developed by the Census Bureau.

our regression procedure is the best way to do it. Freedman and Navidi have focused their critique on the regression method, but they have not addressed the broader question of whether our adjustments improve the census estimates of population distribution.

To illustrate the point, in Section 6, Freedman and Navidi assert that we underestimated the standard errors of our composite estimates by 40%, and thus exaggerated the improvement over the original PEP local area sample estimates. However, even the standard errors of the sample estimates averaged less than 2%. This contrasts sharply with estimated undercounts of 5 or 6% for many areas. For these, the sample and the composite estimates each improve upon the census, and the adjustments are considerably larger than their standard errors. We can improve upon the census. The remaining question is limited to which of several methods of adjustment is best. We have considered four alternatives.

2. FOUR ALTERNATIVE ADJUSTMENT METHODS

The first alternative is to base synthetic adjustments on demographic analysis. If the black undercount rate for the nation is 5% and the nonblack rate is 0%, then a city which is 100% white would get no adjustment, a city which is 20% black would have its population increased by 1%, and a city which is 40% black would increase by 2%. This method has the virtue of relying on demographic analysis which many observers consider to be the most reliable source of information on the undercount. It has drawbacks in that it has no information on Hispanics and does not account for place-to-place variations within the same group. As Table 1 indicates, these matter.

The second alternative is to apply the results of

Table 1. Areas over 40% black or Hispanic would have their populations increased by 5.3%, areas 10–40% black or Hispanic would increase by 1.2%, and all other areas would increase by 0.1%. We decided not to apply this procedure for two reasons. 1) We would be ignoring other factors which might distinguish among areas with similar minority percentages, and 2) our drawing of category boundaries seemed arbitrary. Cities with 39 and 41% black or Hispanic populations, respectively, probably have undercounts that are more similar to each other than to the undercount of a city with a population of 11% black or Hispanic.

The third alternative is to use the sample estimates provided by PEP. There are 66 of these, and Table 2 gives examples for 11 central cities and the 11 states of the Old Confederacy. All 22 areas have substantial black or Hispanic populations, but the cities have higher crime rates, lower mailback rates, and more census-taking problems than the southern states. Each group is fairly homogeneous with respect to these factors, and we would expect undercount rates to be similar within each group. The sample estimates do appear to be internally consistent, although the presence of sampling error inflates the between area differences. There is no apparent reason why South Carolina should have an undercount as high as 5.8% or Tennessee an overcount as high as 2.9%. Regression smoothes these sampling fluctuations, and there is demonstrated precedent of its ability to do so sensibly (Ericksen, 1974; Fay and Herriot, 1979). It was the basis for our fourth alternative.

The regression procedure also allowed us to test our expectations about correlates of the undercount. There were many candidate variables for the regression equation and their strongly positive correlations with the sample estimates of undercount confirmed our theoretical expectations. We settled on three—the crime rate, the percentage of black or Hispanic, and the percentage conventionally enumerated—because they minimized the unexplained variance in the undercount rates. As shown in Table 2, the central cities, with an average sample estimate of 4.9%, had uniformly higher regression estimates than did the southern states, where the average sample estimate was 0.7%. We wanted to allow for the possibility that uniquely local conditions mattered, and we therefore gave some weight to the sample estimates in calculating a final composite. It seems reasonable to suppose that local census-taking problems may have caused an unusually high omission rate in South Carolina or erroneous enumeration rate in Tennessee. The fact that composite estimates have smaller standard errors than the sample estimates indicates that improvements were obtained by systematically borrowing information through regression.

Any of the second, third, or fourth methods would improve the census population estimates by increasing the sizes of large central cities where census-taking problems were concentrated and it is possible that the first (synthetic) method would have done this as well. Improvements would be smaller elsewhere where undercounts were also smaller. We prefer the fourth method because it permits the testing of a prior theory of the undercount.

3. FREEDMAN AND NAVIDI'S ARGUMENT

Freedman and Navidi have taken us to task on a number of points. Many of these are addressed in Kadane's comment or in a longer paper we have coauthored (Ericksen and Kadane, 1985). I would like to respond here to three points: 1) the substitution of

TABLE 2

Sample, regression, and composite estimates of population undercount for 11 central cities and 11 southern states

State	Estimate			City	Estimate		
	Sample	Regression	Composite		Sample	Regression	Composite
		%				%	
South Carolina	5.8	1.2	1.6	Dallas	7.0	4.6	5.0
Louisiana	2.2	1.3	1.4	New York City	6.4	4.0	4.6
Florida	1.6	2.1	1.9	Philadelphia	5.9	2.0	2.5
North Carolina	1.1	0.5	0.7	Baltimore	5.8	4.5	4.7
Mississippi	1.0	0.6	0.7	Los Angeles	5.3	4.0	4.3
Virginia	0.7	0.4	0.5	Houston	4.8	3.1	3.2
Georgia	−0.1	1.1	0.6	Cleveland	4.7	4.1	4.2
Alabama	−0.4	0.8	0.4	Washington D.C.	4.0	5.4	5.2
Texas(R)	−0.5	1.3	0.7	Chicago	3.6	2.9	3.0
Arkansas	−1.0	−0.2	−0.4	Detroit	3.3	5.2	5.1
Tennessee	−2.9	0.1	−0.6	Saint Louis	3.1	6.0	5.7

Notes: Texas(R) refers to that part of Texas remaining after Dallas and Houston have been removed. The sample estimates are provided by PEP series 2/8. The regression estimates are based on an equation where the independent variables are the percentage of black or Hispanic, the crime rate, and the percentage conventionally enumerated. The 11 central cities are all of the cities in the PEP where the percentages of black or Hispanic exceeded 40%.

series 10/8 for series 2/9, 2) the use of the crime rate variable, and 3) the problem of extrapolating to areas other than the 66 for which the regression equation was calculated.

Series 2/9 and 10/8 differ in three important respects. One is that omissions for 2/9 were based on the April CPS, while those for 10/8 are based on the August CPS. The second is that different assumptions were used for 1% of the cases missing data in the sample used to estimate erroneous enumerations. The third, and most important, is that series 10/8 excludes from the analysis anyone who moved between April and August, whether the Census Bureau determined their count/omission status or not. Exclusion of this group, whose omission rate was higher than average, had the effect of reducing national and local undercount rates (see Freedman and Navidi's Table 2). Except for this overall reduction, the differences between the two series are not profound, and the correlation between the two sets of composite estimates is 0.95. This indicates that areas which have particularly high or low undercount estimates on one series have similarly high or low estimates in the other series. When we replaced series 10/8 with 5/8, and thus included movers in the analysis, the overall levels of undercount were similar, and the correlation increased to 0.96. This shows that the effects of replacing April with August data (series 2/5) or changing the missing data assumption for erroneous enumerations (series 8/9) are minimal. We prefer series 2/9 because its assumptions seem more reasonable, but the choice of a PEP series is not crucial.

We relied upon the crime rate variable because including it in the regression equation minimized $\hat{\sigma}^2$. Replacing the crime rate with alternatives such as the percentage of urban, the percentage having difficulty with English, or a dummy variable indicating whether or not the area is a central city, or even deleting the crime rate and using only two variables did not change the composite estimate very much.

Freedman and Navidi make a legitimate point when they state that the crime rate is poorly measured in some areas. In vacation spots like Alpine County, California or Daytona Beach, Florida, the population is swelled by temporary visitors who may be victims or perpetrators of crimes. In such places, the "crime population" is not the same as that the Census Bureau tries to count. Elsewhere, recorded crime rates may be too low if police department records are incomplete. What our results show, however, is that for the 66 areas, in spite of the problems of measurement, the crime rate provided the best fit, indicating that it is a good proxy for the types of census-taking problems encountered in central cities. A decision to replace it with an alternative providing not quite such a good fit, but with fewer errors of measurement, would require subjective judgment which is not outside the statistician's domain. I consider this to be a possible refinement which went beyond the purpose of our testimony in New York.

As Freedman and Navidi show in their Table 4, the problem of using the crime rate variable becomes more serious when the job is to extrapolate beyond the 66 areas for which the regression equation was calculated. Their point is a good one, although their selection of 12 small and unrepresentative counties exaggerates their case. Kadane and I (1985) have replicated their analysis for equal probability samples of 28 central cities, 28 remainders of standard metropolitan statistical areas (SMSA), and 46 nonmetropolitan counties. Substitution of the urbanization for the crime rate variable caused discrepancies in the regression estimates of 3% or more in only 4 of 102 cases, and the discrepancy was less than 1% in 63 cases.

When discrepancies occurred, it was because the crime rates were extreme (e.g., 183 per 1000 population in Daytona Beach and 156 in the Aleutian Islands). The crime rates were more variable in the 102 sample localities than in the 66 sample areas, and there was a weaker relationship between the percentage minority and the crime rate. In a nutshell, the problem was one of extrapolating beyond the values of the independent variables observed in the 66 areas.

This is less of a problem of regression analysis than of PEP design. The 66 areas, which are internally heterogeneous, have much less variation in values of independent variables than do the localities where undercount rate estimates are needed. It would be better if the PEP sample areas were defined in such a way that they better reflected population diversity. We need PEP sample areas with extremely high and extremely low crime rates; other sample areas should have high crime rates and low percentages minority, while still other areas should have low crime rates and high percentages minority. Some PEP areas should be suburban while others should be rural, where the conventional method was used exclusively or not used at all. There is no need to recognize state boundaries in defining such areas. I consider this to be a major design recommendation for the 1990 PEP and contend that it would obviate much of the extrapolation problem that Freedman and Navidi have correctly raised.

4. SUMMARY

When we presented our adjustment procedure in the New York census case, we chose to rely upon the regression method for three reasons: 1) the regression coefficients showed that our estimates were consistent with theoretical expectations, 2) the method provided good, albeit imperfect, measures of uncertainty, and 3) we were able to show that the effects of replacing one PEP series (2/9) with another (5/9) as the dependent variable in regression had only a slight effect on the adjustments.

Freedman and Navidi, who have raised some good points, have properly drawn attention to two important issues: the need to incorporate all sources of error into our measure of uncertainty and problems of extrapolating from the set of areas on which the regression equation is calculated to the set of areas where estimates are needed. These points modify, but do not obviate, the use of our method for adjusting the census. They also fail to demonstrate that our adjustments do not improve upon the census-estimated population distribution for 1980.

An ideal composite estimate would incorporate information from demographic analysis, make allowances for other independent variables that could have been included in the regression equation, and give some weight to alternative series of PEP estimates. Use of the additional sources of information would improve the estimates while increasing our measures of uncertainty. This uncertainty would not increase to the point where we would consider the adjusted population for New York City to be less accurate than the census count. Moreover, the uncertainty associated with our adjustment would be less than the uncertainty with which we must currently view the count.

ADDITIONAL REFERENCES

ERICKSEN, E. P. (1980). Affidavit submitted to U.S. District Court, Southern District of New York, in *Carey v. Baldrige,* 80 Civ., 4550 (HFW).

ERICKSEN, E. P. (1974). A regression method for estimating population changes of local areas. *J. Amer. Statist. Assoc.* **69** 867–875.

ERICKSEN, E. P. and KADANE, J. B. (1985). The robustness of local undercount estimates in the 1980 U.S. Census. Presented at the *International Symposium on Small Area Statistics, Ottawa, Canada, May 24, 1985.*

FAY, R. E., III, and HERRIOT, R. A. (1979). Estimates of income for small places—an application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.* **74** 269–277.

UNITED STATES BUREAU OF THE CENSUS (1982). Coverage of the national population in the 1980 Census by age, sex, and race. *Curr. Population Rep.* Series P-23, No. 115.

Comment

A. P. Dempster

In their provocative article, Freedman and Navidi argue vigorously against the use of "statistical models" for adjustment of 1980 census counts for both large and small regions of the U.S., "even compared to nothing," but indicate that they might allow exceptions if the assumptions were "made explicit" and were "shown to be appropriate." I agree with the authors that explicitness of assumptions is a virtue, but I question whether anyone actually assumes models in so true versus false a form as Freedman and Navidi appear to suggest. Hence, the concept of what is appropriate is considerably more subtle than they allow.

I will discuss below the aspects of modeling which I believe are most critical for regression adjustment of undercount rates derived from the Post Enumeration Program (PEP). I will also take a brief look at the logical foundation of the argument of Freedman and Navidi and I will argue that they have fallen into traps of their own choosing. I agree with them that the frequentist concept of modeling the production of data as "random draws from a box" is only marginally relevant to the applied problem, not only because the methodology is questionable in the specific circumstances, but also and more fundamentally because my attempts to find or construct a satisfying and explicit general account of frequentist logic have all failed. Freedman and Navidi apparently recommend doing "nothing," which I take to be a recommendation to report raw census counts and no more. I prefer a more cheerful outlook. Statistical logic does have merit, and we do have formal tools capable of addressing problems which most professions relegate to guesswork by acknowledged experts. I suggest pushing ahead with a more satisfactory logic. Finally, my comments will conclude with a brief review of the technical development of Freedman and Navidi.

In their zeal to attack certain formal assumptions, Freedman and Navidi risk demolishing statistical principles which lie at the root of our profession's claim to make a scientific contribution to uncertainty assessment. I wish to elaborate on two of these: the principle of randomization and the principle of regression to the mean.

The PEP program does rely on data from formally randomized surveys. The advantage of randomization does not lie primarily in providing a basis for mean square error computations or for randomization tests or confidence intervals, although these may sometimes

A. P. Dempster is Professor of Theoretical Statistics, Department of Statistics, Harvard University, Science Center, One Oxford Street, Cambridge, MA 02138.