imputations in PEP. Last but not least, Ericksen and Kadane have shown courage and innovation by putting forward a methodology in an area fraught with extreme difficulty.

For a broad range of uses the census data are accurate enough, like Newton's laws prior to the discovery of the theory of relativity. A higher intended standard of accuracy, deriving from one man one vote principles and large fund allocations tied to census results, seem to demand a new level of precision. Yet, we have not evolved the needed "theory of relativity" in the area of census adjustment, nor the statistical measuring instruments which could serve as yardsticks when approaching the speed of light. Parenthetically, given the very high level of intercensal mobility and the relatively crude methodology available to track it, it is not entirely obvious why the census must have such extraordinary point-in-time precision. Indeed, over a decade the most disadvantaged areas in terms of congressional representation are undoubtedly those having the highest growth rate.

I am not optimistic about the likelihood of overcoming the technical difficulties involved by 1990, but the issue is clearly important enough so that a major effort must be made.

ADDITIONAL REFERENCE

FELLEGI, I. P. (1981). Should the census be adjusted for allocation purposes? Equity considerations. *Curr. Top. Survey Sampling* 47–76.

Comment

Lincoln E. Moses

This paper shows one side of an argument between two sets of statisticians. The argument was a court case between the country's biggest city and the federal government, with many millions of dollars at stake. No wonder it is fascinating reading. Perhaps it is more surprising that upon reflection I find this paper very convincing, even though I have read just this one side.

Convincing and important.

Freedman and Navidi first describe the census, the Post Enumeration Program (PEP) series, and the approach of New York City to estimating census undercounts by regression of PEP estimates on a number of demographic covariates for 66 areas.

Then they lay bare the assumptions on which depends the validity of the analysis offered by New York City. There are seven such assumptions and the authors give us ample reason to doubt each one. Theorems, real-world heuristics, computations, and experimental sampling are all drawn upon, leaving this reader persuaded that New York City had little claim to having shown a way to improve the census figures by means of regression adjustment.

Freedman and Navidi show that some assumptions are implausible on their face (for example, the independence of two kinds of error component, and that variance of one of them could be regarded as known.)

Lincoln E. Moses is Professor of Statistics in the Departments of Statistics and Family Community and Preventive Medicine of Stanford University. His mailing address is Department of Statistics, Stanford University, Stanford, CA 94305.

They establish that the model entails the assumption that bias in the PEP figures is *not* related to the very demographic variables that are supposed to account for much of the bias in the census, the variables that are to be used to correct the census bias (undercount). They comment on the implausibility of this assumption, and then construct a second series of PEP adjustments, rather parallel to the series used by New York City and find that the difference between the two adjusted series is highly correlated with the demographic variables, which implies that at least one of the two PEP series must fail the key assumption that bias in PEP be unrelated to the demographic variables. The argument to this point implies that biases (assumed away by New York City) are likely operating, making standard errors inadequate measures of error. Then, by means of bootstrap sampling emerges the empirical information that indeed the New York City standard errors (given by formulas appropriate to the theoretical model) do understate the mean square error obtained by empirical sampling from a model in which many of the assumptions by New York City were made true by construction.

Freedman and Navidi have not attacked a strawman, they have not simply set out to find flaws in an example, they have assumed the burden of showing that New York City has not shown how to use the PEP estimates, plus regression, to give improved census counts. If they have succeeded in this (as I think), why is it important to statisticians?

First, statistical argument is becoming more frequent in litigation, so our profession is learning by doing. This case is an instructive example; it shows

that very good statistical work is necessary if it is to carry the day in court, especially under peer review, as practiced in courts.

Second, the nature of the critique is exceedingly instructive, especially if taken as a guide to self-critique. We see plainly the wisdom of stating the assumptions that justify a proposed analysis else someone may state them for us. Then it will be helpful to ponder those assumptions and their implications ("In effect, New York was taking the position that bias in the census—the undercount—was well related to the three explanatory variables, but bias in PEP was not.") Perhaps we can find, as they did, empirical

checks on some of the assumptions. Consideration of several more or less equally plausible alternative models may help us to gauge the fragility of conclusions that we draw from some one of them. We can hope that assessment of error by bootstrap rather than by theoretical formula will become standard practice where experience does not already point to the successful applicability of theoretical error formulas.

These are practical guides to action as self-critic. In my estimation they are the important message in the careful and clearly stated critique given by Freedman and Navidi.

Comment

Gad Nathan

For many years the statistical and legal controversy about the necessity and the advisability of adjusting census counts on the basis of information available from evaluation surveys or from external sources has centered on the general principles involved, such as the definition of the concept of "statistical defensibility" (Spencer (1982) and the discussion thereof). While discussion of these general principles is important, and even necessary, Freedman and Navidi, together with Ericksen and Kadane (1984), are to be congratulated on getting down to the brass tacks of the problems involved in attempts at the real life application of adjustment methods as well as with the theoretical and empirical criticism of the methods proposed. Discussion of actual applications of adjustment methods is important since both proponents and opponents of the adjustment of census counts are generally in agreement that adjustment should be carried out if and only if there exists a method for carrying it out which meets certain conditions and standards of quality and accuracy, for example, by some definition of "statistical defensibility." Since obviously no formal existence or nonexistence theorems can be proved in this respect, the argument must hinge on the empirical results of the proposed methods of adjustment.

Even when an adjustment method can be demonstrated to be adequate, it will not generally be unique and different adjustment methods may be required or may be suitable for the different purposes for which

Gad Nathan is Professor of Statistics in the Department of Statistics, Faculty of Social Sciences, Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel, and is also in charge of Research and Development at the Central Bureau of Statistics, Israel.

census data are used. This raises the issue of who should do the adjusting, the producer of the data or their users. In any case, a fundamental decision to be made with respect to the adjustment procedure involves the definition of the unit of analysis and of prediction, and in particular, the geographical breakdown to be used. Although both Ericksen and Kadane (1984) and the present paper consider different alternative definitions of the geographical areas to be used, they both implicitly assume that the geographical breakdown must be such that reasonably adequate sample estimates of the undercount be available for each area considered.

In fact, the Lindley and Smith (1972) strategy used here has been generalized by Pfeffermann and Nathan (1981) to deal with the important case where observations on the dependent variable are not available for all units, whereas observations on the independent variables are. They also propose methods for estimating the variance σ^2, in this situation, which are similar to those discussed here. Using these results, alternative geographical breakdowns, not necessarily limited to those with sample data for all areas and possibly more suited to the required uses of census data, could be considered.

Most of the criticism of the New York proposals for the adjustment justifiably centers on the underlying assumptions of the regression model and on their justification or lack of it. Indeed, the correct identification of a working model is crucial in this situation where the pure design-based estimate for a single area has too large a sampling error to be of any practical use on its own and must "borrow strength" from other areas via the model-based approach. However, the search for a good model need not be limited to the aim of adjustment via a model-based estimator. A model which can be used not only to estimate more efficiently