

the undercount but also to explain it and to discover the sources of underenumeration can become an important tool for census planners in their attempts to reduce the undercount in future censuses, by attaining a deeper understanding about the underlying mechanism of underenumeration.

As pointed out in the paper by Freedman and Navidi, the proponents of the New York adjustment procedure failed to provide sufficient justification for the model used. This failure was both with respect to the inclusion of variables and the resulting potential bias and with respect to the specification of the variance and of the error structure. A long list of additional potential variables is recommended for consideration. This list includes "geographical location" and interactions, so that the possibility of different regression models for geographical regions, not only with different constants but also with different regression coefficients, must be considered. If we add to this the various possibilities for error structure (model errors, sampling errors, and correlations between them), the number of different models to be considered and the number of their parameters becomes very large indeed. The choice of the correct model among these and the estimation of its parameters all on the basis of 66 observations becomes a formidable problem. To this are added the problems due to the fact that the observations are based on data from a complex sample design, rather than on simple random sampling, so that, for instance, the diagonality of the sampling variance matrix, K , is indeed difficult to justify.

However, in fact, the 66 estimates of undercounts are each based on many observations (the Post Enumeration Program sample size in each area) and this individual information for subunits might be utilized for more efficient model search and identification. For instance, some method of sample re-use or cross-validation based on sample-splitting as proposed by Pfeffermann and Nathan (1985) could be used. It is shown there that efficient cross-validation can overcome both the problem of overfitting and underestimation of error due to the search among a large number of alternatives and the problem of testing goodness of fit on the basis of data from complex samples.

Rejoinder

D. A. Freedman and W. C. Navidi

To begin with, we would like to thank Morrie DeGroot for his editorial support and the discussants for their careful work. We wish Jay Kadane weren't

The empirical results and simulation study of Section 6 illustrate clearly the faults of the proposed adjustment. However, it should be pointed out that the fact that replacement of the crime rate variable by an urbanization rate results in approximately the same quality of fit (as measured under the model assumptions) does not in itself invalidate either model for purposes of adjustment. Similarly, the lack of consistency in the choice of the best subset of three explanatory variables in the simulation study does not necessarily show inadequate adjustment. It is possible that more than a single choice of a set of explanatory variables can provide equally adequate estimates of undercount, although, of course, the explanation provided by the models is thereby limited. In any case, as pointed out, the estimates of standard errors used to judge the quality of these models are definitely deficient.

Finally, although the results of this paper show, without doubt, that the adjustment procedure proposed by New York is not "statistically defensible," this should, under no circumstances, be regarded as a demonstration that an adequate adjustment procedure cannot be found. The negative result should rather be interpreted as implying that an adequate procedure for adjustment of census counts has not yet been found, either for a specific aim or for an official, all purpose one. However, the methods proposed by Ericksen and Kadane (1985) are certainly worthy of further consideration and, above all, for further empirical testing. In particular, suitable methods for model choice and model identification for these circumstances should be developed and applied. The results obtained should be continually scrutinized and appraised by methods similar to those of the present paper.

ADDITIONAL REFERENCES

- PFEFFERMANN, D. and NATHAN, G. (1981). Regression analysis of data from a cluster sample. *J. Amer. Statist. Assoc.* **76** 681-689.
- PFEFFERMANN, D. and NATHAN, G. (1985). Problems in model identification based on data from complex sample surveys. *Bull. Int. Statist. Inst.* **51** 12.2.1-12.2.17.
- SPENCER, B. D. (1982). A note on statistical defensibility. *Amer. Statistician* **36** 208-209.

quite so angry with us, but then we are being very negative about some of his work. He and Gene Ericksen are good statisticians who believe in what they do;