DISCUSSION

Robert E. Fay, U.S. Bureau of the Census

Susan Hinkins describes a survey in which an initial sample of IRS forms is drawn and the entries recorded for each sample case. From this initial sample, matrix sampling is employed to select a subsample of items on forms. Editing of the subsampled items frequently uncovers errors in categorizing income amounts according to definitions intended for a final statistical report. A hot deck imputation procedure is used to impute changes measured on the subsampled items to the rest of the sample. The author discusses use of multiple imputation to measure the effect on reliability of the combined implications of subsampling and imputation.

The application of methods in this paper are quite sound and highly commendable, and these ideas should be of general interest to survey practitioners. Because of the overall appeal of this approach, it appears appropriate to emphasize a specific feature of this application that strengthens its practical utility: that relatively small, although nonetheless important, changes in amounts are being imputed to initial values collected for all sample cases. The fact that the changes are relatively small is critical. If these same techniques were instead applied to collect additional characteristics for which no initial measurements were available, more concern over potential limitations of this approach would be warranted.

In more general applications, the techniques suggested in this paper could encounter substantial difficulties in properly estimating or representing covariance between subsampled characteristics. In the application to the problem in this paper, the initial measurements available for the entire sample could be expected to represent most of the true covariance between items, and the effect of imputation on this covariance would be modest. In more general applications, however, the matrix sampling design might effectively prohibit consideration of one subsampled characteristic in imputing another subsampled characteristic through a hot deck procedure. Much of the true covariance between subsampled characteristics could be lost in such instances.

The author also proposes use of multiple imputation to measure the effect on reliability of subsampling and imputation, but in general applications considerable care would also be required here. Unlike many problems of missing data, the response mechanism, which in this case arises from the matrix sampling design, is known and can be fully incorporated into the choice of adjustment cells for the hot deck. A potential source of difficulty, however, is the scope of missing data, since the subsample may represent only 10 percent or less of the total sample, while this level of response would be intolerable if due to more common sources of missing data, such as refusals. Consequently, this high level of nonresponse extends beyond the more usual level of application of multiple imputation, and properties of this technique for high levels of nonresponse could merit special study.

As a general observation, application of multiple imputation often involves more than simply reapplying the original imputation procedures. For example, many hot decks are designed to avoid, to the extent possible, multiple uses of the same "donor" record by employing, whenever feasible, sampling without replacement to assign a "donor" complete case to an incomplete case requiring imputation. The standard interpretation of multiple imputation effectively assumes, however, sampling with replacement. Thus, other researchers applying these ideas are urged to pay particular attention to the theoretical work of Donald Rubin and others on this methodology.

The preceding cautionary remarks are restricted to hypothetical misapplication of the ideas in Susan Hinkins' paper to other problems. The actual application discussed in her paper is exemplary and illustrates important methodologies in survey research.

The paper of Brian Greenberd and Rita Surdi describes a systematic approach to the relationship between editing and imputation. Their paper provides an important perspective upon problems of missing data by stressing the importance of logical relationships. For example, they point out that in some instances, particularly in the applications to economic surveys and censuses (which have been the principal focus of their efforts), missing data may be reconstructed from logical relationships or known patterns of reporting error. This important point has probably been overlooked by other practitioners on more than one occasion.

The authors discuss a system structured to verify the acceptability of imputations for missing values with respect to the specified edits. The guiding principle for this strategy is to avoid creating, through imputation, data that would have been judged unacceptable in earlier edits, if reported. As a minor point, regression equations for imputation are described whose resulting predicted values require comparison to the edits to insure validity. Perhaps such edits could be incorporated into the form of the equation and the estimation of the coefficients as an alternative solution.

James Harte describes a missing data problem that, although important, is of relatively modest scale by sample survey standards. Because true values become available eventually, this application permits a direct test of the method's performance. The results he presents are fortunately encouraging.

Link relative estimation represents a particular solution to a missing data problem with X_i (past value) observed for all sampled i, and Y_i (current value) observed only for cases in which the response indicator, R_i, takes the value 1. Although the paper emphasizes aspects of the problem unique to link relative estimation, it is also helpful to view this application as an example of the general problem of missing data with respect to the nature of underlying assumptions and effect upon reliability.

Hopefully, the author will be able to take further advantage of his data to investigate the error of prediction at the individual level and whether any response effects might be present.

Nonignorable response mechanisms might include "deaths," where nonresponse may be a consequence of the survey unit having left the population of producers entirely.

S. Earwaker, N. Brien, and J.-F. Gosselin have described promising beginnings in their research at Statistics Canada on small area estimation for labor force characteristics. Some of their empirical findings are similar to those of Gonzalez and Hoza (1978), who studied similar problems in obtaining satisfactory small area estimates of unemployment in the U.S. Both studies indicate that local unemployment is difficult to predict, with relative errors of prediction becoming larger with decreasing size of place.

One alternative investigated by Gonzalez and Hoza was the use of a synthetic estimate as an independent variable in the regression model. Earwaker, Brien, and Gosselin, who investigate properties of the SPREE estimator, could consider a similar use of the SPREE estimator in the regression equation, although the sampling distribution and other properties of this approach would require further study.

As the authors are clearly aware, the effect of sampling variance from a current survey may be quite different on regression and SPREE estimators, and the results presented in their paper do not reflect this consideration. Presumably, this issue will be one of the key questions to be addressed by their future research.

Homer Jones and Paul McMahon review the history of the sample design for the IRS Corporate Income Tax Sample. This paper provides a fascinating illustration of the process by which sample designs become complex. During the lifetime of this survey, the evolution of its design follows a consistent pattern:

a) Stratification in the design has grown elaborate.
b) Increased attention appears to have been given to the question of allocating the sample between strata.
c) The computer has become a progressively important tool in the selection of the sample.
d) The survey estimator has become increasingly complicated.

This increasing complexity in the design led to a significant benefit: the overall sample size has been substantially decreased over time while overall measures of reliability have been maintained. Although the lesson of this history would be anticipated by experienced designers of sample surveys, the paper provides a useful demonstration for those with less experience in this area who may have viewed with skepticism the necessity for the complex methods often encountered in survey designs.

Their paper discusses recent developments in the choice of estimator for this survey. One choice, the post-stratified estimator, computes weights W_{ij} for industry i in sampling stratum j as

$$W_{ij} = N_{ij} / n_{ij}$$

where N_{ij} and n_{ij} represent universe and sample counts in this cell, respectively. The unfortunate liability of this estimator, however, is its instability for cells in which the expected value of n_{ij} is small but variable.

Raking or iterative proportional adjustment has been considered as an alternative estimator for this design. Raking has the anticipate advantage over post-stratification in being less subject to highly variable weights for cells with small expected sample counts. Raking does not fully exploit the available information, however, since its original motivation was for problems in which, for example, population marginals $N_{i.}$ and $N_{.j}$ were known, but not N_{ij}. For the IRS application, however, N_{ij} is available. Consequently, estimators that make some use of the N_{ij} might improve upon raking.

The current estimator, modified raking ratio estimation (Leszcz, Oh, and Scheuren 1983), represents a compromise between raking in its original form and post-stratification, combining the best features of both. In this estimator, post-stratification is employed for all cells with sufficient sample size to support this choice. Raking is used for the matrix of remaining cells, with a constraint on the amount of modification to the inverse probability weights.

In choosing between two estimation alternatives on the basis of sample size, which is itself a random value, the modified raking ratio estimator may be regarded as discontinuous at the cutoff sample size. A possible alternative would be to form a composite estimate combining the two approaches as a smooth function of sample size. This approach might offer slight improvements in precision. Its principle advantage, however, would be to facilitate variance estimation, since linearization or replication is more appropriately applied to some composite estimators than to the modified raking ratio estimator now in use.

Robert Clickner, Glenn Galfond, and Lawrence Thibodeau describe an evaluation of the current sample design for the IRS corporate sample. Their approach is well organized and began with a careful effort to assess the needs of principal users of the data. The authors evaluated the current and alternative designs for meeting these needs.

One observation or conclusion that the authors drew in assessing the needs of users was that some users required precision in the estimates of aggregates while other users required detail. For example, detailed relationships are necessary in order to assess the implications of policy change. This contrast between needs for precision and detail arises for other survey applications, although in practice virtually all survey design is formally evaluated in terms of precision. Thus, the authors' effort to quantify the need for detail concerns a question of wide methodological interest.

Unfortunately, there seem to be few, if any, easy answers to the problem of specifying need for detail. The criteria proposed in the paper, that the percent zero in the sample for specific items closely approximate the population proportion, and that the sample have approximately the same dollar coverages as the population, are not clearly motivated as representations of the needs for detail of survey users. Perhaps the authors will pursue this question further.

The authors compare the current design against two alternatives, each of which uses optimization for the given test data, for example, Neyman allocation. The three designs are then evaluated on the same data. As a technical note, one should always be aware that using the same data for design and evaluation generally guarantees an overly optimistic assessment. This consideration needs to be taken into account in comparing the three designs.

Since the survey has multiple uses, the authors chose to present many of comparisons between designs graphically. Provided that the material is well organized to facilitate visual comparisons of relevant quantities, such an approach may be used to summarize complex relationships between designs. The authors have provided a helpful illustration of this technique.

References

Gonzalez, M.E. and Hoza, C. (1978), "Small Area Estimation with Application to Unemployment and Housing Estimates," Journal of the American Statistical Association, 73, 7-15.

Leszcz, M.R., Oh, H.L. and Scheuren, F.J. (1983), "Modified Raking Estimation in the Corporate SOI Program," Proceedings of the Section on Survey Research Methods 1983, American Statistical Association, Washington, D.C. 434-438.