DISCUSSION

Keith Rust, Westat, Inc.
1650 Research Blvd., Rockville, MD 20850

First, I would like to thank the three speakers for their very insightful and far-reaching papers and presentations. Thanks also to the organizer, Phil Kott, for arranging to have three such worthwhile presentations in the same session. I would also like to express my appreciation to the authors and the chair, Rick Valliant, for providing readable versions of the manuscript well in advance of the session.

Calibration Estimators - Deville and Särndal

The paper provides a generalized approach to developing weight adjustments for survey samples, to reduce the level of sampling error below that achieved using straight Horritz-Thompon weights. The use of such a generalized approach provides two significant strengths that give the potential for significant future practical developments. First, such an approach provides a systematic means of discovering and developing new estimators with desirable properties for specific applications. Second, such a generalized development provides an approach to evaluating comparatively alternative estimators in a given application. The manuscript provides a thorough investigation of the asymptotic properties of calibration estimators.

Following the development given in the manuscript, two questions arise naturally in a given application: 1. What is the best choice of the weights q_k? 2. What is the best choice of distance function G_k; equivalently what is the best choice of the function F? The authors attempt to convert intuitive ideas about what constitute "bad" weights W_k in an application (negative weights, extreme weights), to arrive at an appropriate choice of q_k and G_k. The power of their generalized approach would, I believe, be greatly enhanced if this process could be reversed. If a rationale could be developed for what constitute desirable q_k and G_k, then this would lead naturally to a set of W_k having suitable properties. The W_k would then have a general and objective justification, rather than being defined somewhat on the basis of intuitive ideas about suitable weights. There may in fact be instances when negative weights prove to be desirable, for example.

There are two current practices in survey weighting that might well be improved by applying an extension of the authors' approach. The method of weight trimming, whereby extreme survey weights for individual cases are reduced in an effort to reduce sampling variance, is commonly used but generally ad hoc and lacking a generalized justification. Perhaps the calibration estimator approach can be used to optimize and/or evaluate trimming procedures. Traditional poststratification procedures reduce variance while retaining unbiased estimation. In practice, however, often some bias is actually introduced because, in the notation of the authors, $E(x_{tw}) \neq x_t$. Can the calibration estimator approach be extended to develop a systematic approach to determining when the gains in reduced variance from poststratification outweigh the losses resulting from introduced bias?

With regard to the issue of variance estimation, one useful and general approach is available through the use of replicated variance estimators, such as Jackknife Repeated Replication or Balanced Repeated Replication (see Wolter (1985), Rust (1985)). The advantage of such an approach is that the weight development procedure can be replicated appropriately, and the results "captured" on the survey data file through a set of replicate weights (see Dippo, Fay, and Morganstein (1984)). These weights are then available for secondary data analysts, who can appropriately estimate variances for estimates obtained using calibrated weights, without concerning themselves with the details of the calibration procedure. An example of this approach for a set of iteratively raked survey weights is given in Rust and Bethel (1991).

Finally, I would be interested to learn how this current work relates to the recent advances in using logistic and exponential adjustments of sampling weights, developed by Folsom (1991).

Linear Regression Coefficients - Kott

The author discusses issues of inference about linear regression coefficients. He raises the question as to what are the effective degrees of freedom for the denominator of a given test statistic concerning the regression parameters. This concern applies equally to the derivation of confidence intervals. The issues raised are very important for analysts of clustered or highly differentially weighted data, which frequently result from sample surveys.

The author proposes both a conventional t-test, and also a modified test statistic (T*) with a "less model-biased estimate of the denominator." The two are asymptotically equivalent, so that empirical evaluations of the relative performances of these two estimators would be very informative in assessing which of these estimators is preferable in practice.

Kott uses the technique of the method of moments to derive an appropriate number of degrees of freedom for the distribution of a given test statistic. This general Satterthwaite correction approach provides an intuitively appealing solution to the problem of how to incorporate the effect of the complex design on questions of inference. Two methods are suggested for implementing this approach in practice: 1. Use a formulation involving unknown population quantities, and "guess" at their value, or 2. Substitute sample estimates for these quantities, and estimate the degrees of freedom accordingly. Kott points out the second approach requires the use of fewer assumptions, but my concern is that in many instances this second approach is likely to be substantially biased, especially for multistage stratified designs. This is because variance in the estimation of the population quantities, which translates to bias in estimating the degrees of freedom, is often substantial. Ultimately, a good survey design is one for which the analyst is sure that all values for the denominator degrees of freedom in analyses will be sizeable (30 or more), in which case the exact value of the number of degrees of freedom is not critical.

Finally, I believe that the proposal of adopting a Bonferroni/Simes procedure for testing hypotheses involving multiple parameters is an important one. For adherents of design-based inference, such an approach is probably an essential feature of such analyses, since this is perhaps the only way of adequately accounting for the different degrees of freedom that pertain to different parameters.

Estimation Functions for Interval Estimation - Binder

As with the other papers in this session, this paper proposes a generalized approach to a problem of estimation and inference. In this case, the author has derived a general test-based method of deriving confidence intervals for model parameters. This approach is developed to give a general method of deriving confidence intervals for a subset of a set of multidimensional parameters.

The approach of using test-based confidence intervals has great intuitive appeal. Why then has such an approach not been more widely used heretofore? Two reasons are apparent. The first is that such intervals are generally much less straightforward to compute than those based on a large sample approximation. The second is that, as reported by Cochran (1977, Section 6.5), there is a suggestion that, for the ratio estimator, such an approach is not adequately conservative. Consequently, I think that empirical evaluation of the test-based approach for common survey applications will be a necessary component for their widespread acceptance in practice.

In considering the merits of test-based confidence intervals compared to those obtained traditionally using large sample assumptions, it is important to step back and consider more generally what constitute desirable properties for two-sided confidence intervals. As I see it, there are five desirable properties for a method of obtaining confidence intervals: 1. The interval should have the correct coverage properties (or be conservative); 2. the interval should have small expected length; 3. The interval should have small variance in length; 4. the interval should be symmetric about the point estimate; 5. the tail distributions should be symmetric, that is, the interval should be as likely to be too high as to be too low.

Traditional confidence intervals of the form $\hat{x} \pm t.\ \widehat{se}(\hat{x})$ satisfy 4), but can perform very poorly with regard to 5) (see for example Kovar, Rao, and Wu (1988)). Yet I would argue that in fact 5) is a much more desirable property to possess than 4), given a choice among procedures roughly equivalent with regard to properties 1) - 3), since I believe that most users of two-sided confidence intervals derived from survey data implicitly assume property 5) even when it is far from true (whereas this is not the case for property 4)). Intuitively, it seems possible that test-based confidence intervals might, in many cases, be superior to the traditional large sample asymptotic confidence with regard to property 5). If this can be demonstrated, I believe that this will be a powerful argument for their wide adoption, in addition to the case made by Binder.

References

Cochran, W.G. (1977). Sampling Techniques. Third edition. Wiley: New York.

Dippo, C.S., Fay, R.E., and Morganstein, D.H. (1984). Computing variances from complex samples with replicate weights. Proceedings of the Section on Survey Research Methods, pp. 489-494, American Statistical Association: Alexandria, VA.

Folsom, R.E. (1991). Exponential and rogistic weight adjustments for sampling and nonresponse error reduction. Proceedings of the Section on Survey Research Methods. American Statistical Association: Alexandria, VA.

Kovar, J.G., Rao, J.N.K., and Wu, C.F.J. (1988). Bootstrap and other methods to measure errors in survey estimates. Canadian Journal of Statistics, Vol. 16, Supplement, pp 25-45.

Rust, K.F. (1985). Variance estimation for complex estimators in sample surveys. Journal of Official Statistics, 1, 381-397.

Rust, K., and Bethel, J. (1991). An experiment within a survey: the design of the 1990 Trial State Assessment of Educational Progress. Proceedings of the Section on Survey Research Methods. American Statistical Association: Alexandria, VA.

Wolter, K. M. (1985). Introduction to Variance Estimation. Springer-Verlag: New York.