Journal of Official Statistics, Vol. 20, No. 3, 2004, pp. 441–443

Discussion

Robert M. Groves[1]

This lecture is especially well-suited to its venue because Hansen and his colleagues initiated a manner of thinking about survey measurement errors and offered statistical models that are the foundation of much of our current practice. The naming of Paul Biemer as this year's Hansen lecturer is appropriate as he, in my opinion, is his generation's major contributor to our understanding of response variance in sample surveys.

My comments on the lecture begin with a short history, setting a slightly different context than that offered by Biemer. Given that context I want to focus on four main features of work on simple response variance: a) measurement, inference, and reduction, b) implicit and explicit models, c) sensitivity analysis for model-based assumptions, and d) bias versus variance.

1. History

The Hansen, Hurwitz, and Pritzker (1964) paper developed a basic framework that motivated the use of reinterview measurements as a tool to estimate simple response variance for key statistics. Such reinterviews were mounted in an attempt to capture a component of measurement error that was seen to vary over conceptual trials of the measurement, *given the same measurement conditions*. Encouraged by the developments of Hansen et al. and, I suspect, influenced by his will, the U.S. Census Bureau developed a program of reinterviews that exists to this day. It is notable that, while some other central statistical bureaus around the world may have similar programs, the practice of using reinterview measurements to estimate measurement error variance remains a rarity. While probability sampling as a tool to measure sampling variances clearly became the dominant practice for scientific uses of the survey method, similar adoption did not largely occur with designs permitting measurement of simple response variance.

2. Measurement, Inference, and Reduction

One problem of the simple response variance models is that they employ untestable assumptions, given the usual practice of reinterview studies. Further, there are two clear issues that arise when reinterview measurements are proposed. One is the cost of mounting a recontact of a sample unit. In these days of declining response rates, survey budgets are under fierce pressures, and the marginal costs of reinterview studies are not as easily justified when the initial measurements themselves are threatened. Another is that the statistical culture of surveys and the field culture of data collection have built-in conflicts.

[1] University of Michigan, Survey Research Center, P O Box 1248, Ann Arbor, MD 48106-1248, U.S.A. Email: bgroves@isr.umich.edu

It *has* become standard practice to run verification reinterviews for at least a portion of respondents. The field operations purpose for these is largely to verify that no fabrication of interviews has occurred, and, to a lesser degree, evaluate the performance of the interviewer. For these goals, higher-level staff are generally used (potentially producing more inconsistent interviewing behavior than originally), modes of data collection cheaper than that originally used are often chosen (potentially producing more inconsistent measurement conditions), and the length of time between the first and second measurement often is variable (potentially producing the risk of change in true values between the two measurements).

The probability sampling inferential paradigm, I suspect, became so dominant in sample surveys because it offered not only a strategy to *measure* an error source (sampling) but also a set of tools to *reduce* the error (stratification and differential assignment of selection probabilities given prior information about the frame population). The original work in simple response variance was governed by a components of variance perspective, without guidance on what features of the design might be altered to affect the magnitudes of the variance components. Thus, while there are undoubtedly file drawers of simple response variance estimates on U.S. Census Bureau surveys over the past forty years, it is a legitimate question to ask whether they have led to improved survey measurement and estimation. My guess is that the track record is quite spotty.

First, I know of few practical uses of simple response variance estimates at the inferential step in sample survey estimation. While they have been constructed to estimate a variance component that is additional to that arising from sampling, they, by and large, lay unused at the inference stage. That is a failure.

Second, the response variance models do not suggest how to reduce simple response variance. Indeed, most errors in surveys have been reduced through identification of the cause of the error, and the models offer no guidance on this. It is noteworthy that the application of social and cognitive psychological concepts to surveys has produced large advances in our understanding of the causes of response errors. Thus, the field has learned how to reduce response errors through choice of mode of data collection, question wording and question structure. These developments, however, arose almost independently of empirical results on simple response variance. Clearly, one remaining challenge is to integrate the statistical models of simple response variance with the causal models of response errors. Such an integration could combine measurement with reduction efforts in a useful feedback of continuous improvement.

3. Implicit and Explicit Models

Many of the simple response variance approaches assume independent trials or unconditional independence in order to estimate the variability in response deviations. Unfortunately, practical measurement situations rarely offer independence possibilities. Human respondents remember past measurements; they change over time. The resulting covariance between measurements obtained on two trials can bias most parameters. The deadly feature of the assumption is that there are generally no empirical ways to assess whether it is valid. Biemer notes that latent class models offer a "local independence" assumption, whereby the two trials are independent, conditional on a true value. However, it is not at all clear what practical difference this assumption makes for estimation.

As with all models, there is real value in having enough flexibility in the parameterization that sensitivity analysis of conclusions based on different parameterization is possible. The credibility of empirical estimates rises when the conclusions based on them remain the same when alternative plausible assumptions are made. Running many versions of somewhat flawed models can provide great insight. Here, Biemer's praise of the Hui-Walter sensitivity analysis in a latent class framework is well placed.

4. Bias versus Variance

The Biemer article focuses on binary variables, which present a confound between bias and variance properties. It uses the Hansen, Hurwitz, Pritzker index of inconsistency as the focus. It may be preferable (following the progress made in cognitive models of response formation) to first specify the mechanisms of response error that are producing the phenomenon of variability and then deduce a statistical model that mirrors that phenomenon. I join with Biemer in his attraction to some of the approach of Guggenmoos-Holzman on this score.

5. Summary

Biemer has offered us a good summary of decades of development in simple response variance and has alerted us to developments outside sample survey statistics. This is valuable. He has also alerted us to missed opportunities of integrating these models into the inferential step of survey estimation and minimizing survey errors. There is much left to do.

Received February 2004