Journal of Official Statistics, Vol. 18, No. 2, 2002, pp. 155–156

Discussion

Chris Skinner[1]

This article provides an excellent overview of uses of models in survey statistics and demonstrates Graham Kalton's great ability to elucidate issues of survey methodology, which matter in practice. I shall only comment on general aspects of the article.

A central theme is the view that design-based inference is to be preferred as a basic approach, but that in practice it has a variety of limitations, which models can play a useful role in fixing up. My remarks are broadly in line with this view in the case of descriptive inference but not in the case of analytic inference.

For descriptive inference, the design-based and model-based approaches provide a principal contrast. Here, in the absence of the many practical problems considered by Graham, Neyman's (1934) basic case in favour of the design-based approach seems compelling. Design-based inference statements (possibly conditioned as Graham indicates) are valid under plausible assumptions, meet many practical needs and have a clear interpretation, despite concerns that these statements do not meet all desirable criteria (Valliant et al. 2000). On the other hand, the validity of model-based inference statements will typically depend upon less plausible assumptions regarding either a model or balance with respect to arbitrary control variables.

Of course, the many practical issues covered in this article and related considerations, such as outliers and influential observations, may make it either difficult or inappropriate to sustain a purely design-based approach. Indeed, the combined focus in the article on both practice and the design-based approach should not be interpreted as implying that design-based methods are always more practical than model-based methods. The objective that estimation be design consistent wherever possible has been a theoretical ideal, one which appears to have been maintained in official statistics to a large extent because advocates such as Morris Hansen have emphasised its importance. In other fields of survey practice, such as areas of market research with which I am familiar, inference appears to meet practical imperatives without the use of probability sampling or design-based inference.

An important benefit of a statistical modelling framework is that it provides clear principles for deriving efficient point estimators of specified parameters. As Graham describes, the model-assisted approach aims to achieve a comparable benefit for design-based methods by mapping models to estimators. A problem is that this mapping can be rather arbitrary or ad hoc. Optimality properties do not necessarily translate. For example an optimal model-based regression predictor is not necessarily mapped onto an optimal

[1] University of Southampton, Department of Social Statistics, Highfield, Southampton SO17 1BJ, U.K. Email: cjs@socsci.soton.ac.uk

design-based generalized regression estimator (Rao 1994). Further methods for mapping model selection or diagnostic procedures onto methods for choosing design-based estimators are not well developed (Nascimento Silva and Skinner 1997).

Error is a central concept in survey methodology and I think that the value of statistical modelling as a framework for assessing error also merits emphasis. One illustration given by Graham is the use of models for variance estimation with non-measurable probability sampling designs. More generally, I think models can play an essential role in assessing estimation error arising from a range of sampling and non-sampling sources. The assessment of error under alternative assumptions, such as in a sensitivity analysis, seems to me to be an essential characteristic of a modelling approach, in contrast to a design-based approach which aims to be ''assumption-free.'' For example, Kokic (1998) estimates the variance of a complex economic index using a parametric bootstrap approach under a series of alternative modelling assumptions. Of course, such a modelling framework may still leave a place for design-based methods. For example Kokic (1998) uses design-based methods to estimate some of the variance components required to feed into the bootstrap approach.

For analytic uses of survey data, where model parameters are of interest, I find it more natural to view model-dependent inference as the preferred basic approach, and then to consider the possibility of using design-dependent methods to fix up this inference, if there are grounds for supposing that the sample selection is non-ignorable. Model building, including the use of diagnostic techniques, seems usually to be a more scientifically relevant way to specify parameters than via finite population considerations, except in ''near-descriptive'' set-ups. Design-dependent weighting methods may, however, help to fix up problems of bias from non-ignorable sample selection, and design-based methods may also be useful for variance estimation when the variance effect of the design is non-ignorable.

References

Kokic, P.N. (1998). Estimating the Sampling Variance of the U.K. Index of Production. Journal of Official Statistics, 14, 163–179.

Nascimento Silva, P.L.D. and Skinner, C.J. (1997). Variable Selection for Regression Estimation in Finite Populations. Survey Methodology, 23, 23–32.

Rao, J.N.K. (1994). Estimating Totals and Distribution Functions Using Auxiliary Information at the Estimation Stage. Journal of Official Statistics, 10, 133-152.

Received March 2002