Journal of Official Statistics, Vol. 28, No. 3, 2012, pp. 353–357

Discussion

Risto Lehtonen[1]

1. Introduction

In his inspiring article, Professor Little presents a fresh view of statistical inference by extending the scope of calibrated Bayes introduced in the author's previous papers of 2006 and 2011. Little proposes calibrated Bayes as an alternative inferential framework – even a new inferential paradigm – for official statistics. The article supplements the debate on important theoretical foundations and paradigms of statistical inference nicely. Good examples of recent discussion are Rao (2011) and Särndal (2010), as well as the collection of papers published in the Pakistan Journal of Statistics (Special Issue to mark the 80th Birthday of K.R.W. Brewer, 2011).

If a model-assisted framework is taken as the prevailing paradigm in official statistics (which Little does, at least under the design-based framework), I realize – in a serious application of Thomas Kuhn's classical theory of paradigm shifts in science – that it is not fruitful to discuss new theories using arguments taken solely from the prevailing paradigm. Moreover, history shows that the actual paradigm shifts only can be verified afterwards! Little's article treats several additional aspects of interest. I restrict my discussion to three specific topics related to inference in official statistics: the role of administrative registers in official statistics production, inference in the context of official statistics, and a discussion on innovation in official statistics.

2. The Role of Administrative Registers in Official Statistics Production

Little describes the status quo for statistical inference at the U.S. Bureau of Census as a combination of design-based and model-based ideas (design/model compromise, DMC). He believes that DMC constitutes the current inferential philosophy in official statistics agencies in general. This obviously is true for statistical infrastructures where official statistics are mainly produced using sample surveys, and administrative registers play a secondary role. But this does not necessarily hold for "register countries", that is, countries where the statistical infrastructure relies strongly on unit-level administrative registers that successfully cover the whole target population (here, administrative data refer to any data collected primarily for some purpose other than official statistics production). Professor Little only devotes a brief discussion (e.g., Section 4.8) to the role of administrative data in official statistics production. Let me address this point in more detail.

[1]University of Helsinki, Department of Social Research, P.O. Box 33 (Gustaf Hällströmin katu 2b), FI-00014 Helsinki, Finland. Email: risto.lehtonen@helsinki.fi

There are an increasing number of statistical institutes, notably in Europe (Denmark, Finland, Norway and Sweden as forerunners), where a large share of official statistics is produced using micro-merged statistical registers. The combination of different administrative data sources into integrated statistical registers at the unit level is based on identifiers that are unique to the various data sources. Personal identification numbers and business identification codes are good examples of such identifiers. For example, at Statistics Finland an increasing share of official statistics produced using register information, covering areas of social, business and environmental statistics and regular register-based population censuses. In Finland, this option is based on the Finnish Statistical Act. It contains regulations concerning the different stages of statistics compilation and focuses especially on the rationalization of data collections, data supplier relations, implementation of statistical ethics and data protection (Statistics Finland 2007). For example, regional and small area statistics at detailed aggregation levels are often produced using statistical registers. Because of the obvious benefits (high cost-effectiveness, low respondent burden, small share of missing data, lack of measurement error due to respondent etc.), many national statistical agencies are successfully progressing towards a more extensive use of administrative data in statistics production. Perhaps the option of administrative registers also is relevant for the U.S. Census Bureau, but this aspect remains untouched in the article by Professor Little.

There also is a clear connection of register statistics with inferential issues relevant for official statistics: accuracy assessment and inference under register-based survey statistics are in their infancy, calling for more research input on the subject (recent examples can be found, for example, in papers presented in the European Conference on Quality in Official Statistics Q2012). The calibrated Bayes approach proposed by Professor Little might offer alternative and new views of inference under register-based information, but this option perhaps was purposely left out of the scope of the article.

3. Inference in the Context of Official Statistics

Prof. Little accords the accounting for survey design features a central role in calibrated Bayes inference. As stated in the article, all inferences under the calibrated Bayes approach are Bayesian and hence model-based, but models are constructed to yield inferences with good design-based properties such as design consistency. In the estimation of finite population parameters of interest, design consistency is obtained through the inclusion of survey design information in the estimation procedure. Design consistency can be achieved by the inclusion of survey design variables in the model, which is a standard model-based routine in protection against model misspecification. Alternatively, design weights can be incorporated in the estimation of the model (this also is proposed by Little, see Eq. (11)). This technique is commonly employed in design-based model-assisted estimation. Design consistency has received much attention in the literature of model-based methods and in hierarchical Bayes small area estimation in particular (e.g., You and Rao 2003; Lahiri and Mukherjee 2007; Datta 2009). A classical example in this context is the pseudo EBLUP (empirical best linear unbiased predictor) presented by Rao (2003).

Professor Little argues that "design consistency", not "design bias", is the important issue, since the essence of shrinkage estimates is that exact unbiasedness is secondary to mean squared error, and for possibly design-inconsistent estimators, design bias is not an important component of mean squared error (Section 4.3). He also states that design consistency is a rather weak property (Section 4.5). Obviously, design consistency tends to lose power as a quality criterion when turning to small sample situations or the estimation of subpopulation quantities, for example. In my view, design bias still is an important issue for official statistics production. It has been shown that model-based estimators such as EBLUP type estimators of totals for population subgroups or domains (small or large) can be severely design-biased, but they can have small variance (e.g., Lehtonen, Särndal and Veijanen 2003, 2005). The bias can dominate the MSE under model failure, and in this case, valid design-based confidence intervals of an estimator of totals for one or several population subgroups are not necessarily obtained. Similar observations have been made for more complex statistics such as at-risk of poverty rate, estimated by EBP (empirical best predictor) and logistic GREG (generalized regression) type estimators in small area estimation with underlying logistic mixed models (Lehtonen et al. 2011). In my opinion, the problem of possible design bias still deserves more attention in the context of calibrated Bayes estimation.

For Professor Little, model-assisted estimators represent a rather ad hoc way of making a design-based estimator robust to model misspecification, whereas a more direct approach is simply to choose a more robust model (Section 4.5). Let me address this point in more detail. It is well known that design-based model-assisted GREG estimators of totals for population subgroups (small or large) are nearly design-unbiased by their construction principle, and this property holds even under model failure (Särndal, Swensson and Wretman 1992; Lehtonen and Veijanen 2009). It also should be pointed out that in GREG estimation for population subgroups or domains, the possible population heterogeneity can be accounted for by postulating, for example, a mixed model with domain-specific random effects (e.g., Lehtonen, Särndal and Veijanen 2003). I just wonder if this approach fulfils Little's requirement of "a more robust model". To me, Example 4 of Section 4.5 represents an overly simplified counter-example.

It is fair to say that the price to be paid for the near-unbiasedness of a GREG estimator is the possibly increased variance, which can become visible in small sample situations in particular. There thus is a trade-off between bias and variance, and this dilemma is recognized, for example, by survey statisticians in statistical agencies. Here, the concept of design/model compromise introduced by Little becomes relevant: use of model-based methods for small area estimation is often a better motivated choice, even at the cost of possible design bias.

4. Innovation in Official Statistics

The article of Professor Little motivates an excursion to the role of innovation in the production of official statistics. Professor Little treats this aspect to some extent in the concluding section and I agree with many of his points. Experiences from many statistical institutes show that the prevailing cultural environment in an agency plays a central role when trying to introduce new inferential approaches and methodological innovations into statistical production processes. The background is the following. The main routine task in a typical statistical institute is to deliver reliable statistics with good accuracy for finite population characteristics, such as monthly unemployment in the country and its subregions. In this work, the agency struggles with the trade-off between bias and variance. Under the prevailing official statistics culture, the main focus tends to be on obtaining small design bias, even at the cost of increased variance. Statistical agencies are aware that ignorable bias can be achieved by design-based methods such as model-assisted estimation, which is a widespread approach and is generally considered as a well-established inferential framework for official statistics production. Thus, official statistics culture tends to be conservative rather than innovative with respect to inferential approaches and practices, and it is difficult to implement any significant changes. Cultural difficulties in implementing new inferential or methodological approaches for statistics production in national statistical agencies are discussed in more detail for example in Lehtonen and Särndal (2009) and Platek and Särndal (2001), and earlier in Dillman (1996).

Now, let me take a more optimistic position with respect to innovation spread. The results of Lehtonen and Särndal (2009) suggest that the introduction of new ideas and approaches for official statistics production can be more successful in agencies that interact closely and permanently with academic research compared to agencies whose cooperation with the world of universities is less active. There are an increasing number of agencies in the former group and they share properties such as a strong R&D infrastructure within the agency and well-established cooperational arrangements with university departments, jointly funded professorships and joint research projects. Under these circumstances, the introduction and testing of inferential innovations in official statistics might be an option. I strongly agree with Professor Little's proposal of a strong research program within government statistical agencies, including cooperative ties with statistics departments in academic institutions.

In addition to calibrated Bayes, other alternatives have been proposed for innovation in official statistics inference. Good examples of recent keywords are: model-free calibration (Särndal 2007), model calibration (Wu and Sitter 2001), randomization-assisted model-based inference (Kott 2005), and combined survey sampling inference (Brewer 2002). The existence of complementary and competing approaches to inference in official statistics is of course favorable and suggests material for comparative research initiatives, probably to be implemented by consortia of interested university departments and advanced statistical agencies. For example, the forthcoming Horizon 2020 programme (The 8th EU Framework Programme for Research and Development) might offer a possible platform for such initiatives.

5. References

Brewer, K.R.W. (2002). Combined Survey Sampling Inference: Weighing Basu's Elephants. London: Arnold.

Dillman, D.A. (1996). Why Innovation is Difficult in Government Surveys. Journal of Official Statistics, 12, 113–124.

Kott, P.S. (2005). Randomization-Assisted Model-Based Survey Sampling. Journal of Statistical Planning and Inference, 129, 263–277.

Lahiri, P. and Mukherjee, K. (2007). Hierarchical Bayes Estimation of Small Area Means under Generalized Linear Models and Design Consistency. Annals of Statistics, 35, 724–737.

Lehtonen, R. and Särndal, C.-E. (2009). Research and Development in Official Statistics and Scientific Co-operation with Universities: a Follow-Up Study. Journal of Official Statistics, 25, 467–482.

Lehtonen, R., Särndal, C.-E., and Veijanen, A. (2003). The Effect of Model Choice in Estimation for Domains, Including Small Domains. Survey Methodology, 29, 33–44.

Lehtonen, R., Särndal, C.-E., and Veijanen, A. (2005). Does the Model Matter? Comparing Model-Assisted and Model-Dependent Estimators of Class Frequencies for Domains. Statistics in Transition, 7, 649–673.

Lehtonen, R. and Veijanen, A. (2009). Design-Based Methods of Estimation for Domains and Small Areas. Chapter 31 in Handbook of Statistics, Vol. 29B. Sample Surveys. Inference and Analysis. C.R. Rao and D. Pfeffermann (eds). Amsterdam: Elsevier, 219–249.

Lehtonen, R., Veijanen, A., Myrskylä, M., and Valaste, M. (2011). Small Area Estimation of Indicators on Poverty and Social Exclusion. Research Project Report WP2 (D2.2, FP7-SSH-2007-217322 AMELI). Available at: http://www.uni-trier.de/index.php?id=24676&L=2

Platek, R. and Särndal, C.-E. (2001). Can a Statistician Deliver? Journal of Official Statistics, 17, 1–20.

Rao, J.N.K. (2003). Small Area Estimation. Hoboken, NJ: John Wiley & Sons.

Rao, J.N.K. (2011). Impact of Frequentist and Bayesian Methods on Survey Sampling Practice: A Selective Appraisal. Statistical Science, 26, 240–256.

Statistics Finland (2007). Quality Guidelines for Official Statistics. 2nd Revised Edition. Helsinki: Statistics Finland. Available at: http://www.tilastokeskus.fi/meta/qg_2ed_en.pdf

Särndal, C.-E. (2007). The Calibration Approach in Survey Theory and Practice. Survey Methodology, 33, 99–119.

Särndal, C.-E. (2010). Models in Survey Sampling. Official Statistics – Methodology and Applications in Honour of Daniel Thorburn. M. Carlson, H. Nyquist, and M. Villani (eds). Department of Statistics, Stockholm University, 15–27.

Särndal, C.E., Swensson, B., and Wretman, J.H. (1992). Model Assisted Survey Sampling. New York: Springer.

Wu, C. and Sitter, R.R. (2001). A Model-Calibration Approach to Using Complete Auxiliary Information from Survey Data. Journal of the American Statistical Association, 96, 185–193.

You, Y. and Rao, J.N.K. (2003). Pseudo Hierarchical Bayes Small Area Estimation Combining Unit Level Models and Survey Weights. Journal of Statistical Planning and Inference, 111, 197–208.