Journal of Official Statistics, Vol. 28, No. 4, 2012, pp. 471–476

Letter to the Editor

Letters to the Editor will be confined to discussion of articles which have appeared in the Journal of Official Statistics and of important issues facing the statistical community.

Revisiting the Multipurpose Property of Sampling Weights

The recent article by Professor R.J.A. Little (Little 2012) includes a discussion of alternative basic philosophies of official statistics production. In this letter, we wish to bring to the attention of JOS readers another, related fundamental property of significance for official statistics production, which we believe is related to the matter that Prof. Little discusses.

Use of sampling weights is a feature that probably distinguishes survey sampling most from other statistical disciplines. In survey practice, statisticians have traditionally called for them to satisfy the so-called *multipurpose property* (Särndal 2007), that is, that a single set of sampling weights is used to estimate all population variables in a multipurpose survey.

Another key concept in official statistics production is auxiliary information. It occurs at different stages: the sampling design (Cochran 1977), the construction of estimators (Särndal et al. 1992), the treatment of nonresponse (Särndal and Lundström 2005), the imputation methods (Haziza 2009), to name perhaps the most noteworthy. Auxiliary information in statistical offices is nowadays abundant, available, up-to-date and of good quality for statistical purposes.

We contend that this increasing availability of auxiliary information invites us to consider putting aside the multipurpose property. We reason as follows. From a purely theoretical standpoint, there is no reasoning that supports the multipurpose property. Moreover, adhering to methodological rigour in sampling weights construction, one can easily find reasons not to have a single set of weights. Let us consider, for instance, nonresponse treatment. Reweighting for nonresponse (see e.g., Särndal and Lundström 2005; Bethlehem et al. 2011) is an elaborate technique where either calibrating against benchmark auxiliary information or modelling response propensity (also using auxiliary information) assists in the weight adjustment and bias reduction. Regarding calibrating, to take a specific example, the following statement by Ranalli (2008) is enlightening in this respect:

> The calibration approach of Deville and Särndal (1992) has been referred as to be "model-free" (Särndal 2007), as opposed to regression estimation in which an assisting model has to be specified to conduct estimation. We believe that model-free, in this case, refers to being free from an *explicit* [original italics] modelling procedure. In fact, the results reported here show that calibration, although developed in a purely design based framework, *implicitly assumes a linear relationship between all the survey variables and the auxiliary ones* [our italics].

Thus, if the auxiliary information needed to adequately deal with the nonresponse differs between different variables of interest, why would one not use the correspondingly different sets of sampling weights for each of them? Furthermore, if accepting different sets of sampling weights in a multipurpose survey, why not use more accurate techniques, such as, for instance, model calibration (Wu and Sitter 2001; Wu 2003; Montanari and Ranalli 2003, 2005) in the construction of estimators? Moreover, what if we use model-assisted techniques with non-linear models rendering the concept of sampling weight itself surpassed by a more complex, although possibly more accurate, notion of (non-linear) sampling estimator (Lehtonen and Veijanen 1998)? Accuracy is clearly an argument in favour of having several sets of sampling weights.

In the present multidimensional reading of data quality, not using a single set of weights can also be viewed as a possible cost reduction in terms of sampling sizes: if for a given sample size n and its corresponding cost $c(n)$, lower variances, say, $V_{\text{no multi}} < V_{\text{multi}}$ can be achieved by dropping the multipurpose property, why not think of keeping the same accuracy V_{multi} but reducing the sampling size $n' < n$ and the corresponding cost $c(n') < c(n)$? Another example going in the same direction stems from the use of multiple frames with Hartley-Fuller-Burmeister-type estimators. In this case, cost reduction because of the use of multiple frames is also present, but is often disregarded because of the multipurpose property (Lohr 2009).

In a more general discourse, dropping the multipurpose property can be viewed as a chance to use model-based techniques in the construction of sampling estimators. It gives the statistician the opportunity to resort to the vast field of classical inference statistical techniques *without crossing the red line between the design-based and model-based approaches* (see e.g., Smith 1994 and references therein for a detailed discussion). As prominent examples, let us cite model-assisted estimation (Särndal et al. 1992) and model calibration (Wu and Sitter 2001; Wu 2003; Montanari and Ranalli 2003, 2005): the estimators obtained thereby are (approximately) design-unbiased, being protected against model-breakdowns. Typically, they are also more accurate than those not using these statistical-modelling assisting techniques. But the door is open: why not use more general techniques, for instance, geostatistical techniques or time-series modelling, in the same fashion?

On the other hand, from a practical point of view, we can suggest several reasons supporting the multipurpose property, namely, (i) sampling weights interpreted in a sense of representativity, which apparently reinforces the multipurpose property; (ii) the numerical consistency among all output tables in multipurpose surveys; and (iii) the concerns about transparency in data dissemination.

As we see it, the representativity interpretation and the multipurpose property are strongly reinforcing each other in survey practice: if a sampling weight of a sample unit k is interpreted as the number of population units represented by k, it is natural to have just a single set of sampling weights in a survey; and vice versa, if only a single set of sampling weights is to be accepted in a survey, it is natural to interpret them as a measure of the representativity of each sample unit. In our opinion, the representativity view has already been challenged by consequently adopting the theoretically correct interpretation of a sampling weight ω_{ks} in a linear estimator $\hat{Y}_U = \sum_{k \in s} \omega_{ks} y_k$ as a multiplicative factor of the variable value y_k of unit k in the sample s when estimating the population total $\sum_{k \in U} y_k$.

Table 1. Estimates of population units exhibiting and not exhibiting habit A

Habit A	Male	Female	Total
Present	$\hat{Y}_m^A$	$\hat{Y}_f^A$	$\hat{Y}_m^A + \hat{Y}_f^A$
Absent	$\hat{Y}_m^{\neg A}$	$\hat{Y}_f^{\neg A}$	$\hat{Y}_m^{\neg A} + \hat{Y}_f^{\neg A}$
Total	$\hat{Y}_m^A + \hat{Y}_m^{\neg A}$	$\hat{Y}_f^A + \hat{Y}_f^{\neg A}$	.

Notice that negative weights and weights $\omega_k < 1$ are indisputable in this interpretation. The possibility of having negative sampling weights underlines the essential difference between the design-based and the model-based approaches to inference. Without going into detail beyond the scope of this letter, take as an example the issue whether sampling weights must be used in analysing survey data with heteroskedastic linear regression models or not (see e.g., Little 2004 and references therein). Choosing a model variance $\Sigma = \text{diag}\{\sigma_1^2, \ldots, \sigma_N^2\}$, with $\sigma_k^2 \propto \omega_{ks}$, is clearly impossible in the case of at least one negative sampling weight $\omega_{k*s} < 0$. We believe that this is a direct consequence of the irreconcilable difference between the two approaches (Smith 1994). In our opinion, official statistics must remain on the safe side of design-unbiasedness, although model-assisted. However, the notion of a sampling weight as a measure of the representativity of the associated unit should be exorcised from survey sampling (Kruskal and Mosteller 1979a,b,c, 1980).

More importantly, numerical consistency among all output tables in a multipurpose survey is a concern. It is indeed a very serious concern, giving justification to the multipurpose property: it ensures numerical consistency. Let us consider an example in a health survey where the presence or absence of two habits A and B is measured in the population. Let us accept that different sets of weights $\{\omega_k^A\}$ and $\{\omega_k^B\}$ are used because different auxiliary variables have been used in the calibrating stage. Suppose that the results are demanded broken down by sex. This is usually presented in the form of contingency tables as in Tables 1 and 2.

Here the issue becomes apparent: rarely, under the assumed working hypotheses, will the pairs of marginal estimated sex totals $(\hat{Y}_m^A + \hat{Y}_m^{\neg A}, \hat{Y}_m^B + \hat{Y}_m^{\neg B})$ and $(\hat{Y}_f^A + \hat{Y}_f^{\neg A}, \hat{Y}_f^B + \hat{Y}_f^{\neg B})$ coincide respectively. This is the consistency alluded to above. In the sphere of official statistics, this can be very difficult to accept from the point of view of users of the statistics: how is it possible that we can be faced with different male and female counts as a result of estimating different variables? Should these counts not be the same irrespective of the estimated variable?

Thirdly, transparency in official statistics entails anonymised microdata released to final users in such a way that they can reproduce almost any published estimate. The case of

Table 2. Estimates of population units exhibiting and not exhibiting habit B

Habit B	Male	Female	Total
Present	$\hat{Y}_m^B$	$\hat{Y}_f^B$	$\hat{Y}_m^B + \hat{Y}_f^B$
Absent	$\hat{Y}_m^{\neg B}$	$\hat{Y}_f^{\neg B}$	$\hat{Y}_m^{\neg B} + \hat{Y}_f^{\neg B}$
Total	$\hat{Y}_m^B + \hat{Y}_m^{\neg B}$	$\hat{Y}_f^B + \hat{Y}_f^{\neg B}$	.

several sets of sampling weights, or even of estimators assisted with possibly non-linear models, renders this task much more complex, to the point of even preventing the user from computing any further estimate not contained in published releases. Such a lack of transparency could damage official statistics.

Any attempt to drop the multipurpose property in official statistics production must in our opinion tackle all these questions. Firstly, the question regarding the interpretation of sampling weights has already been settled in the methodological arena, but the idea of representativity must be carefully dealt with when disseminating official statistics. Secondly, the numerical consistency of any planned or unplanned table must be guaranteed. That is, there must exist a methodological solution to the numerically consistent estimation of population quantities arranged in almost any cross-tabulation of variables. This is the case both for those tabulations contained in the survey production plan and for those not included but later called for. In this sense, it seems nowadays advisable to move the focus of the problem of estimation in a finite population from its traditional univariate setting (see e.g., Hanurav 1966) to a more general and realistic definition: given a finite population U of known size N and composed of identifiable units with variable values $\mathbf{y}_k$, the objective is to produce numerically consistent estimates for any planned or unplanned set of tables of population quantities $f_p(\mathbf{y}_1, \ldots, \mathbf{y}_N)$, $p = 1, \ldots, P$. In this regard, let us cite the repeated weighting technique (Kroese et al. 2000). Repeated weighting resorts to an extensive use of calibrating provided "one is willing to abandon the common practice of using one set of [. . .] weights [. . .]" (Boonstra et al. 2003). Thus, important steps have already been taken in this direction, although more work needs to be done to reach a satisfactory solution.

To sum up, dropping the multipurpose property arises as an attractive invitation to use statistical models and more general techniques in assisting the construction of survey estimators within the design-based framework. In official statistics, this would pave the way not only for the stimulation of novel ideas on how to adapt these techniques in the construction of estimators, but also the inclusion of existing methods in the daily production of statistical offices in a general fashion. However, we also believe that in official statistics any step in this direction must guarantee the accessibility and clarity of the published information, which must be released in an understandable, suitable and convenient manner to the final user. In current user-oriented statistical systems, we are convinced that some pedagogical actions regarding the chosen methodology and dissemination policies should be considered in order to take into account users' needs and to guarantee maximum transparency.

References

Bethlehem, J., Cobben, F., and Schouten, B. (2011). Handbook of Nonresponse in Household Surveys. Hoboken, NJ: Wiley.

Boonstra, H.J.H., van der Brakel, J.A., Knottnerus, P., Nieuwenbroek, N.J., and Renssen, R.H. (2003). A Strategy to Obtain Consistency Among Tables of Survey Estimates. Workpackage 7 of DACSEIS project. Available from: http://www.dacseis.de. Accessed November, 5th, 2012.

Cochran, W.G. (1977). Sampling Techniques, (3rd ed.). New York: Wiley.

Deville, J.C. and Särndal, C.-E. (1992). Calibration Estimators in Survey Sampling. Journal of the American Statistical Association, 87, 376–382.

Hanurav, T.V. (1966). Some Aspects of Unified Sampling Theory. Sankhya A, 28, 175–204.

Haziza, D. (2009). Imputation and Inference in the Presence of Missing Data. In Sample Surveys: Design, Methods and Applications, D. Pfefferman and C.R. Rao (eds). Amsterdam: North-Holland.

Kroese, B., Renssen, R.H., and Trijssenaar, M. (2000). Weighting or Imputation: Constructing a Consistent Set of Estimates Based on Data from Different Sources, In Statistics Netherlands (2000), Netherlands Official Statistics, 15, special issue on Integrating administrative registers and household surveys. Voorburg/Heerlen: Statistics Netherlands.

Kruskal, W. and Mosteller, F. (1979a). Representative Sampling. I: Non-scientific Literature. International Statistical Review, 47, 13–24.

Kruskal, W. and Mosteller, F. (1979b). Representative Sampling. II: Scientific Literature, Excluding Statistics. International Statistical Review, 47, 111–127.

Kruskal, W. and Mosteller, F. (1979c). Representative Sampling. III: The Current Statistical Literature. International Statistical Review, 47, 245–265.

Kruskal, W. and Mosteller, F. (1980). Representative Sampling, IV: The History of the Concept in Statistics, 1895–1939. International Statistical Review, 48, 169–195.

Lehtonen, R. and Veijanen, A. (1998). Logistic Generalized Regression Estimators. Survey Methodology, 24, 51–55.

Little, R.J.A. (2004). To Model or not to Model? Competing Modes of Inference for Finite Population Sampling. Journal of the American Statistical Association, 99, 546–556.

Little, R.J.A. (2012). Calibrated Bayes, an Alternative Inferential Paradigm for Official Statistics. Journal of Official Statistics, 28, 309–334.

Lohr, S. (2009). Multiple-frame Surveys. In Sample Surveys: Design, Methods and Applications, D. Pfefferman and C.R. Rao (eds). Amsterdam: North-Holland.

Montanari, G.E. and Ranalli, M.G. (2003). On Calibration Methods for the Design-based Finite Population Inferences. Bulletin of the International Statistical Institute, 54th session, vol. LX, contributed papers, book 2, 81–82.

Montanari, G.E. and Ranalli, M.G. (2005). Nonparametric Model-calibration Estimation in Survey Sampling. Journal of the American Statistical Association, 100, 1429–1442.

Ranalli, M.G. (2008). Recent developments in calibration estimation. Proc. XLIV Meeting of the Italian Statistical Society, pages 355–362.

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

Särndal, C.-E. and Lundström, S. (2005). Estimation in Surveys with Nonresponse. Chichester: Wiley.

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

Smith, T.M.F. (1994). Sample Surveys 1975–1990: An Age of Reconciliation? International Statistical Review, 62, 5–19.

Wu, C. (2003). Optimal Calibration Estimators in Survey Sampling. Biometrika, 90, 937–951.

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.

D. Salgado[1]
C. Pérez-Arriero[2]
M. Herrador[2]
I. Arbués[1]

[1] National Statistical Institute, D.G. Methodology,
Quality and Information and Communications Technologies,
Paseo de la Castellana, 28071 Madrid, Spain
E-mail: david.salgado.fernandez@ine.es;
ignacio.arbues.lombardia@ine.es

[2] National Statistical Institute, S.G. Sampling and Data Collection,
Paseo de la Castellana, 28071 Madrid, Spain
E-mail: carlos.perez.arriero@ine.es;
monserrat.herrador.cansado@ine.es