Journal of Official Statistics, Vol. 23, No. 4, 2007, pp. 463–465

# Discussion

*Chris Skinner*[1]

Two main types of application of multiple imputation (MI) to official statistics have been proposed:

1. the use of MI by a statistical agency in its “primary” activity of producing estimates from survey data;
2. the construction of multiply imputed datasets for “secondary” analysis by different users, including the construction of synthetic datasets to protect confidentiality (Raghunathan et al. 2003).

Jan Bjørnstad’s article focuses on the first use and I shall restrict my comments to this topic also. See e.g., Meng (1994) on some of the issues arising with the second use.

The condition for MI to lead to valid inference in a non-Bayesian framework is termed “proper” by Rubin (1987, Chapter 4). I think it is important to recognize that this condition applies not to the imputation method alone, but to the imputation method for a given “complete data” point estimator and variance estimator (i.e., $\hat{\theta}$ and $\hat{V}(y)$ in the notation here). Thus an agency could not necessarily make the standard MI approach valid by drawing imputations in a Bayesian way from the predictive distribution of the missing values, even if this were practical, since such imputation might not be proper with respect to the agency’s methods of point and variance estimation, and thus might lead to biased variance estimation (Wang and Robins 1998; Nielsen 2003; Kim et al. 2006). Instead, the agency would either need to determine an imputation procedure which was proper with respect to its estimation methods, which Binder and Sun (1996) have shown is often extremely difficult for the kinds of methods used by agencies in practice, or would have to consider revising its (complete data) point and variance estimation procedures to fit in with the imputation procedure, which might be viewed as the “tail wagging the dog.”

Jan Bjørnstad’s article considers to what extent the validity of MI might be retained through the less extreme option of modifying the MI variance estimator while retaining the agency’s preferred (complete data) point and variance estimator and imputation method. While one could consider the problem of estimating the variance of an MI point estimator for a nonproper imputation scheme from first principles, as in Wang and Robins (1998), Jan Bjørnstad explores instead the possibility of making a simple modification to the standard MI combination formula via the use of the $k$ term in Expression (1).

[1]University of Southampton, Southampton Statistical Sciences Research Institute, Highfield, Southampton SO17 1BJ, UK. Email: c.j.skinner@soton.ac.uk

The article provides interesting demonstrations, for a number of specific cases, that consistent variance estimation can be achieved by taking $k$ to be a measure of the proportion of missing information, such as the reciprocal of an item response rate. For the approach to provide a principled basis for general applications, in line with the aims of MI, it is desirable to understand the method's potential generality, as discussed in Section 6 and summarised in the Theorem. Two basic conditions of the Theorem seem to me reasonably uncontentious. A number of authors have considered combinations of estimators and imputation methods which obey $E(\hat{\theta}^{*}|y,s) = \hat{\theta}$ (cf. Rubin 1987, Equation 4.2.5; Binder and Sun 1996, Equation 14; Kim et al. 2006, Condition C3) to restrict attention to cases where the MI point estimator is unbiased. Likewise, Condition (3) corresponds to common assumptions (cf. Rubin 1987, Equation 4.2.8; Binder and Sun 1996, Equation 19; Kim et al. 2006, Equation 3.4).

However, in a number of other respects, the conditions (a)–(g) used in A.5 of the Appendix to prove the Theorem seem rather restrictive. Firstly, the MCAR condition (or conditions a and b) has very strong consequences, especially in preventing consideration of estimators with unequal weights (see the Lemma in Section 6) or domain estimators, whereas Kim et al. (2006) show that these are particularly important features of official statistics applications which may lead to bias in the multiple imputation variance estimator. Further restrictions in the Appendix are that: the result seems restricted to a limited class of sampling schemes (for example multi-stage sampling is not discussed) to a limited range of imputation schemes (excluding for example most versions of nearest neighbour imputation methods widely used in official statistics) and to statistics with imputed values in only one variable.

There seem to me therefore to be a number of dimensions of generality that it would be useful to research further. Moreover, to assess the extent to which the proposed approach may be useful in practice, it seems necessary to assess the relative merits of this approach as compared with alternative methods for variance estimation with imputed data such as those of Rao and Shao (1992), Shao and Steel (1999), and Kim and Fuller (2004). Criteria for comparison include the breadth of conditions under which the approaches are valid and the extent to which the methods provide unified approaches for sets of estimands, such as means, totals or proportions, across different domains and different variables. Other standard criteria are efficiency of point and variance estimation and ease of computation.

## 1. References

Binder, D.A. and Sun, W. (1996). Frequency Valid Multiple Imputation for Surveys with a Complex Design. Proceedings of the American Statistical Association, Section of Survey Research Methods, 281–286.

Kim, J.K., Brick, J.M., Fuller, W.A. and Kalton, G. (2006). On the Bias of the Multiple-Imputation Estimator in Survey Sampling. Journal of the Royal Statistical Society, Series B, 68, 509–521.

Kim, J.K. and Fuller, W.A. (2004). Fractional Hot Deck Imputation, Biometrika, 91, 559–578.

Meng, X.-L. (1994). Multiple Imputation Inferences with Uncongenial Sources of Input (with Discussion). Statistical Science, 9, 538–573.

Nielsen, S.F. (2003). Proper and Improper Multiple Imputation. International Statistical Review, 61, 317–337.

Raghunathan, T.E., Reiter, J.P. and Rubin, D.B. (2003). Multiple Imputation for Statistical Disclosure Limitation. Journal of Official Statistics, 19, 1–16.

Rao, J.N.K. and Shao, J. (1992). Jackknife Variance Estimation with Survey Data under Hot-deck Imputation. Biometrika, 79, 811–822.

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: Wiley.

Shao, J. and Steel, P. (1999). Variance Estimation for Imputed Survey Data with Non-negligible Sampling Fractions. Journal of the American Statistical Association, 94, 254–265.

Wang, N. and Robins, J.M. (1998). Large-sample Theory for Parametric Multiple Imputation Procedures. Biometrika, 85, 935–948.

Received June 2006