

Discussion

Ralf T. Münnich¹

1. Variance Estimation in the Presence of Nonresponse

Professor Bjørnstad addresses a new approach to an extremely challenging topic, variance estimation in the presence of nonresponse and imputation. In general, when single imputation in order to compensate for nonresponse is applied, the classical variance estimation techniques tend to underestimate the *true* sampling variance which takes the imputation process into consideration.

Starting with the compensation of nonresponse biases, three approaches may help to reduce this bias when estimating means, totals, or proportions:

- applying weighting techniques, e.g., from the class of calibration estimators (cf. Deville and Särndal 1992 or Demnati and Rao 2004);
- single imputation, which is widely used in National Statistical Institutes (NSIs);
- multiple imputation.

Weighting techniques focus on unit nonresponse problems. In the case of item nonresponse with complicated nonresponse patterns, in general imputation methods are used. The imputation algorithm in connection with the set of auxiliary variables aims at compensating for a possible nonresponse bias assuming missing completely at random or missing at random.

In order to estimate standard errors or to deliver confidence intervals, correct variance estimates have to be elaborated. In the case of imputed data these suffer from underestimating the true variance due to ignoring the randomness of the imputation process. The variance of an estimator $\hat{\theta}(y^*)$ on the imputed data set y^* can be estimated via variance decomposition

$$V = V(E(\hat{\theta}^* | y^*)) + E(V(\hat{\theta}^* | y^*)) \quad (1)$$

which is conditioned on the imputed data set (cf. Berger et al., 2004, p. 7). Ignoring the imputation would yield the inner variance of the latter term which generally leads to a severe underestimation of the true variance of $\hat{\theta}^*$.

When applying single imputation and resampling variance estimation, the variance inflation can be assured by sophisticated weighting methods or applying a single imputation in each resample (cf. Rao 2005; Berger and Rao 2006; Berger and Skinner 2005; Berger et al. 2004). However, these methods are very computer intensive and hardly applicable in large samples in the near future.

¹University of Trier – Faculty IV – Economic and Social Statistics Department, Universitätsring 15, D-54286 Trier, Germany. Email: muennich@uni-trier.de

Another approach was introduced by Rubin (1978), multiple imputation. Applying the so-called Rubin's combination formula one can derive the true variance of the estimate in terms of a variance decomposition taking the mean variance of the estimates and the variation of the point estimates under the different imputed data sets, the inner and outer variance. The advantage of multiple imputation is that a wide range of estimators may be used assuming a variance estimator is available for the statistic. However, proper imputation methods are rarely used in official statistics. Hot-deck methods are generally nonproper and hence cannot be applied in a *classical* multiple imputation framework in the Rubin sense.

Bjørnstad aims at filling the gap between applying classical single imputation methods and multiple imputation methods in order to draw correct inferences from point estimators with imputed data. This is achieved by deriving an inflation factor k within Rubin's combination formula in the following way:

$$V(\hat{\theta}_{\text{MI}}) = \frac{1}{m} \sum_{j=1}^m \hat{V}(\hat{\theta}_j^*) + \left(k + \frac{1}{m}\right) \cdot \frac{\sum_{j=1}^m (\hat{\theta}_j^* - \hat{\theta}_{\text{MI}}^*)^2}{m-1} \quad (2)$$

where $\hat{\theta}_{(j)}^*$ denotes the estimate for the j th imputed data set.

Let me address two important items of Bjørnstad's work, both important for its success in follow-up work and practice:

- The computational effort may become severe in large-scale surveys. Applying the Rao and Shao (1992) bootstrap variance estimator needs at least 100 bootstrap replicates each with a newly imputed data set. The effort of applying Bjørnstad's multiple imputation routine would reduce the computational burden by at least 70% assuming the use of $m = 30$ multiply imputed data sets. The complexity of the reduction is linear. The comparison with jackknife variance estimates is more sophisticated due to the variety of newly developed routines that are based on specialized weighting techniques.
- The generality of the multiple imputation framework of Rubin's work allows the end-user to apply different estimation techniques to the same m multiply imputed data sets. Within Bjørnstad's work, the critical question is related to the variance inflation constant k whether it can be applied to different estimators simultaneously or whether it has to be derived separately for each estimator.

Bjørnstad gives some ideas on how to elaborate the constant k . Under the given assumptions, one may start using

$$k = \frac{1}{1 - \frac{n - n_r}{n}} \quad (3)$$

The major question arises whether this formula holds for practical purposes under

1. the given examples in Bjørnstad's work and
2. beyond these examples under more general imputation rules and other estimators, e.g., the highly nonlinear Laeken-indicators drawn from the European survey on income and living conditions (EU-SILC, cf. Dennis and Guio 2004).

Equation (5) in Bjørnstad gives an idea on how to elaborate estimates for k . However, in order to estimate k from the sample, one has to evaluate the two conditional terms $V(E(\hat{\theta}^*|y_{\text{obs}}))$ and $E(V(\hat{\theta}^*|y_{\text{obs}}))$. The major problem may arise in evaluating the conditioning on y_{obs} . Applying resampling methods may spoil the gain in efficiency mentioned before.

The two problems described above will be examined in a short simulation example given below which is related to the general simulation study conducted within the DACSEIS research project (cf. <http://www.dacseis.de>).

2. An Example

The aim of the small simulation study is to elaborate the critical variance inflation constant k in a practical environment. This example strongly follows the Monte-Carlo study in Münnich and Rässler (2005), whereas the framework of the Monte-Carlo studies in the DACSEIS context is best described in Münnich et al. (2004) and its summary results in Davison et al. (2004).

A classical problem in official statistics is the estimation of the number of unemployed people. In Germany, several definitions of unemployment are to be considered. The jobless, close to the ILO concept, are estimated from the German Microcensus. However, in Germany people register as unemployed in order to ask for subsidies. Within the German Microcensus, a link variable to this register variable is included in the questionnaire, which allows this variable to be used as an auxiliary variable (cf. Wiegert and Münnich 2004). The German Microcensus itself is a 1% stratified cluster sample of households and individuals.

The synthetic universe is based on the DACSEIS universe from the federal state Saarland including three subpopulations:

SUB0 The estimation variable is the number of jobless in the federal state Saarland. As auxiliary variable, the number of people registered as unemployed was taken.

SUB1 The same estimation task was conducted on a subpopulation which is a regional stratum of approximately one third of the size. Due to the smaller size, one may expect heterogeneities to play a more important role than in the case of SUB0.

SUB5 In this subpopulation only larger buildings on the same estimation programme were considered. The population is more homogeneous than SUB0 and SUB1.

This notation is strongly related to the DACSEIS simulation study (cf. <http://rpm.dacseis.de>).

Within the simulation study 10,000 Monte-Carlo samples were drawn to elaborate point and variance estimates with $m = 30$ imputed data sets for each estimator and Monte-Carlo sample. In the case of SUB0 only 1,000 replicates were conducted due to the long computation time of already two weeks on a 3 GHz Intel Pentium PC on Windows XP and the statistical software R. This fact will cause a little lack in precision of the Monte-Carlo variance of the point estimator in connection with the task SUB0. The settings of the Monte-Carlo study are strongly connected to the settings in Münnich and Rässler (2005).

As single imputation routines two hot-deck methods were applied called SI LFS2 and SI LFS3, respectively, to have coherent notation to the DACSEIS simulation study (cf. Laaksonen et al. 2004). These two single imputation routines were applied in the

non-Bayesian multiple imputation framework with $m = 30$ data sets. Further, the multiple imputation routine PRIMA from Münnich and Rässler (2005) was considered, also with $m = 30$ data sets. Both the Horvitz-Thompson (HT) and the generalized regression estimator (GREG), were used to estimate the number of jobless people. Due to the fact that nonresponse was only implemented in the estimation variable, the classical calibration estimator could be taken as a benchmark.

Figure 1 gives an overview of the estimation results for the three different populations. As expected, the GREG estimator and hence the calibration estimator yields the best results and dominates the HT estimates. This is due to the highly correlated auxiliary variable which was used for the calibration and the imputation. One can also observe that the population SUB1 turns out to be very problematic, which results from the heterogeneous small strata to be considered.

The main question in this context is the critical evaluation of the variance inflation factor k . Starting with Bjørnstad's recommendation, one may assume the inflation constant k to be

$$k = \frac{1}{1 - \frac{n - n_r}{n}} = \frac{4}{3} \tag{4}$$

ignoring the stratification. The latter may play a more considerable role for the scenario SUB1 where the strata with respect to house size class are all included and hence more heterogeneous. The internal and external variances can be estimated from the simulations. Further, with the help of Equation (5) in Bjørnstad's article, an optimal constant k_{opt} can be derived, which is shown in Table 1 for the different scenarios.

As expected, the population SUB1 spoils all results. Nevertheless, the quick solution (4) seems to work in the case of the GREG estimators. Applying the HT estimator, one may

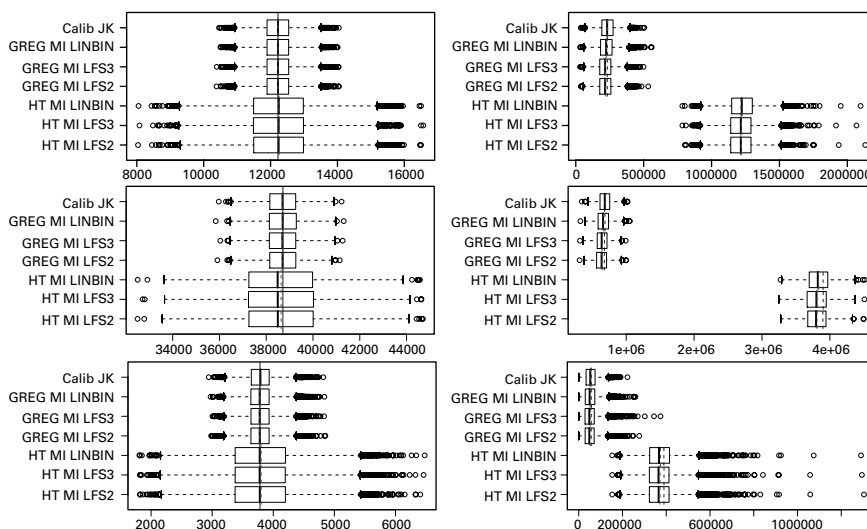


Fig. 1. Boxplots of the point (left) and variance (right) estimation distributions for the three scenarios SUB0 (top), SUB1 (middle) and SUB5 (bottom)

Table 1. Optimal constants k_{opt} for the HT (left) and GREG (right) estimator under multiple imputation with SI LFS2 and SI LFS 3 as well as under MI PRIMA

Estimator Population Imputation method	HT			GREG		
	SUB0	SUB1	SUB5	SUB0	SUB1	SUB5
	$k_{opt,0}$	$k_{opt,1}$	$k_{opt,5}$	$k_{opt,0}$	$k_{opt,1}$	$k_{opt,5}$
SI LFS2	1.6683	2.6946	0.9327	1.2572	1.4292	1.2817
SI LFS3	1.6908	2.6950	0.9526	1.2500	1.3978	1.2930
MI PRIMA	1.4426	2.2696	0.8139	1.1281	1.2169	1.0978

assume a violation of the assumptions given by Bjørnstad. Looking a little into detail, one can find very little sensitivity of the constant k to the results – which in this case is the unbiasedness of the variance estimator under imputation. The latter can be drawn from Figure 2.

The kernel density estimates were drawn from the standard settings within the software R. One can observe a noticeably higher ratio for the HT than for the GREG estimator. Hence, the sensitivity of the variance estimates is much lower for the HT than for the GREG. This results in less biased variance estimates than expected.

The elaboration of the constant k as well as the estimates is not significantly influenced by the number of imputed data sets m assuming this number to be at least 5. The difference between $m = 5$ and $m = 30$ in the given simulation experiment was under 3%. Nevertheless, for robustness reasons and for confidence interval coverage rates under normality one may prefer $m = 30$.

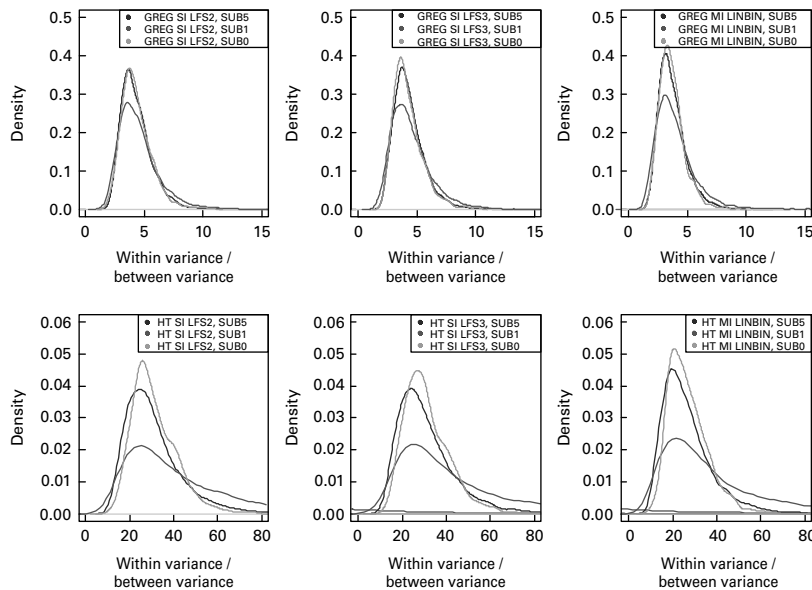


Fig. 2. Kernel density estimates of the ratio of within and between variances within the simulation runs for the three imputation methods SI LFS 2 (left), SI LFS 3 (middle) and MI PRIMA (right) and the two estimators (GREG: top, HT: bottom)

3. Final Remarks

The given example proved within the context of the DACSEIS study to be problematic in some cases, especially under SUB1. Therefore, the simulation results can be seen in the framework of a very unfriendly environment. In the case of the GREG estimates, the results seem to be very good. However, the HT estimates remain a little problematic and in the SUB1 example unacceptable. The latter problem may be reduced by considering separately estimated constants k – the approach of predetermined constants via response rates seems to have space for improvements here.

Further research may enable end-users to apply the methodology themselves also in difficult environments. However, the questions remain:

1. How can one estimate k ideally – at least in cases where the plug-in value (4) may be inappropriate?
2. How robust are the estimation results in connection with bad estimates of k ?

Professor Bjørnstad presented his ideas at the DACSEIS final conference at the Q2004 Conference in Mainz (cf. <http://www.dacseis.de>). His presentation opened a vivid discussion including Professor Rubin as the originator of multiple imputation in the Bayesian context. Personally, I hope that this new approach will lead to further discussions on its applicability and allow wider comparisons between the different above-mentioned methods. One major goal will be the derivation of appropriate estimates of the variance inflation constant k – this may very well bring the two multiple imputation ideas closer to each other and also foster their applicability in practice.

Personally, I would like to congratulate Professor Bjørnstad for his utmost inspiring idea and the eloquent presentation of his article.

4. References

- Berger, Y.G., Bjørnstad, J., Zhang, L.-C., and Skinner, C.J. (2004). Imputation and Nonresponse, DACSEIS – Workpackage 11, Deliverable 11.1.
- Berger, Y.G. and Rao, J.N.K. (2006). Adjusted Jackknife for Imputation under Unequal Probability Sampling without Replacement. *Journal of the Royal Statistical Society, Series B*, 68, 531–547.
- Berger, Y.G. and Skinner, C.J. (2005). A Jackknife Variance Estimator for Unequal Probability Sampling. *Journal of the Royal Statistical Society, Series B*, 67, 79–89.
- Davison, A., Knottnerus, P., Münnich, R., Ollila, P., and Skinner, C.J. (2004). The DACSEIS Recommended Practice Manual, DACSEIS -Workpackage 12, Deliverables 12.2 and 12.3.
- Demnati, A. and Rao, J.N.K. (2004). Linearization Variance Estimators for Survey Data. *Survey Methodology*, 30, 17–26.
- Dennis, I. and Guio, A.-C. (2004). Poverty and Social Exclusion in the EU. *Statistics in Focus 16/2004*, Eurostat, Luxembourg.
- Deville, J.-C. and Särndal, C.-E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, 87, 376–382.
- Laaksonen, S., Oetliker, U., Rässler, S., and Skinner, C.J. (2004). Imputation and Non-Response, DACSEIS – Workpackage 11, Deliverable 11.2.

- Münnich, R., Boonstra, H.J., Eckmair, D., Magg, K., Oetliker, U., Quatember, A., Renfer, J.-P., Sardy, S., and Wagner, H. (2004). Variance Estimation in Complex Surveys. DACSEIS-Workpackage1, Deliverable 1.2.
- Münnich, R. and Rässler, S. (2005). PRIMA: A New Multiple Imputation Procedure for Binary Variables. *Journal of Official Statistics*, 21, 325–341.
- Rao, J.N.K. (2005). Re-sampling Variance Estimation with Imputed Survey Data: Overview. Institute, I.S. (Ed.) Proceedings of the 55th Session of the International Statistical Institute, Sydney.
- 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. (1978). Multiple Imputation in Sample Surveys – A Phenomenological Bayesian Approach to Nonresponse. Proceedings of the American Statistical Association, Section of Survey Research Methods, 20–40.
- Wiegert, R. and Münnich, R. (2004). German Register Data for Regression Estimation in Survey Sampling – A Study on the German Microcensus Respecting for Data Protection. *Jahrbücher für Nationalökonomie und Statistik*, 224, 247–259.

Received January 2007