Resampling Variance Estimation in Surveys with Missing Data

A. C. Davison and S. Sardy*

February 6, 2006

Abstract

We discuss variance estimation by resampling in surveys in which data are missing. We derive a formula for jackknife linearization in the case of calibrated estimation with deterministic regression imputation, and compare the resulting variance estimates with balanced repeated replication with and without grouping, the bootstrap, the block jackknife, and multiple imputation, for simulated data based on the Swiss Household Budget Survey. Jackknife linearisation, the bootstrap, and multiple imputation perform best in terms of relative bias and mean square error.

Some key words: Balanced repeated replication; Bootstrap; Calibration; Influence function; Jackknife; Linearization; Missing data; Multiple imputation.

*Sylvain Sardy is a postdoctoral researcher and Anthony Davison is Professor of Statistics, both at Institute of Mathematics, School of Basic Sciences, Ecole Polytechnique Fédérale de Lausanne, Station 8, CH-1015 Lausanne, Switzerland, (http://stat.epfl.ch). This work was performed in the context of the European Union project DACSEIS (http://www.dacseis.ch). We thank the other members of the DACSEIS team for their valuable collaboration.

1 Introduction

Classical variance formulae for sample survey estimators are derived using approximations based on Taylor series expansion of the estimators. When the sample is small or the estimator complex—for instance, because of modifications to account for missing data—it is natural to be concerned about the quality of such approximations, and to consider alternatives such as resampling procedures. The purpose of this paper is to give formulae for general variance approximations in the presence of calibration and deterministic imputation, and to compare them numerically with resampling procedures.

Section 2 reviews the classes of estimator that we consider, and Section 3 reviews resampling methods for variance estimation. Section 4 outlines a jackknife linearisation approach for use when missing data are dealt with by calibration and deterministic regression imputation, and Section 5 gives a brief numerical comparison based on the Swiss Household Budget Survey. The paper ends with a brief discussion.

2 Basic ideas

Consider first complete response for a stratified single stage unequal probability sampling scheme without replacement, with N units divided into H strata, from which a total of n units are sampled. Let n_h be the number of units sampled from the N_h population units in stratum h, and let π_{hi} be the inclusion probability for unit i of this stratum. In household surveys this unit might consist of a cluster of individuals, in which case the unit response of interest is supposed to be cumulated over the cluster. Let x_{hi} and y_{hi} be variables that have been measured on the units, where y_{hi} is the scalar response of interest and x_{hi} is a $q \times 1$ vector of auxiliary variables, which may be continuous, categorical, or both.

Parameters of the finite population can be classified into two broad groups. The first, largest, and most important group comprises smooth quantities such as the population total $\tau = \sum_{h=1}^{H} \sum_{i=1}^{N_h} y_{hi}$, the ratio, the correlation, or the change in the ratio between two sampling occasions. The other main group comprises non-smooth functions of the finite population responses, such as the median, quantiles, and statistics based on them (Berger and Skinner, 2003).

Estimation of the finite population parameters is based on the data from the n sampled units and on their inclusion probabilities under the given sampling design. The most important estimator of a total is the Horvitz–Thompson estimator

$$\widehat{\tau} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \omega_{hi} y_{hi} = \omega^{\mathrm{T}} y, \qquad (1)$$

where $\omega_{hi} = 1/\pi_{hi}$ are the inverse inclusion probabilities. The variance of $\widehat{\tau}$ is readily obtained, but complications arise when the weights themselves are random, or when some of the responses are unavailable.

In many cases population totals are known for some of the auxiliary variables x, and this information can be used to increase precision of estimation. Suppose that q_C marginals of the q auxiliary variables are known, with $q_C \leq q$, let c be the $q_C \times 1$ vector of known marginals, and let X_C denote the $n \times q$ matrix of auxiliary variables whose marginal total for the entire population is known to equal c. Using the estimation of a total to illustrate the idea of calibration, the quality of the Horvitz–Thompson estimator can be improved by choosing the weights w_{hi} to be as close as possible to the original weights ω in some metric G, subject to the constraint that the weighted auxiliary variables match the marginals (Deville and Särndal, 1992), that is,

$$\min_{w_{hi}} \sum_{h=1}^{H} \sum_{i=1}^{n_h} \omega_{hi} G(w_{hi}/\omega_{hi}) \quad \text{such that} \quad \sum_{h=1}^{H} \sum_{i=1}^{n_h} w_{hi} x_{Chi} = c.$$

A distance measure widely used in practice is the ℓ_2 or squared error metric, $G(x) = (x-1)^2/2$; then the calibrated weights equal

$$w = \omega + \Omega X_C (X_C^{\mathrm{T}} \Omega X_C)^{-1} (c - X_C^{\mathrm{T}} \omega), \tag{2}$$

where Ω denotes the diagonal matrix whose elements are the ω_{hi}. So the calibrated Horvitz–Thompson estimator of the total is

$$\begin{aligned} \widehat{\tau} &= w^{\mathrm{T}} y = \omega^{\mathrm{T}} y + (c - X_C^{\mathrm{T}} \omega)^{\mathrm{T}} (X_C^{\mathrm{T}} \Omega X_C)^{-1} X_C^{\mathrm{T}} \Omega y \\ &= \omega^{\mathrm{T}} y + (c - X_C^{\mathrm{T}} \omega)^{\mathrm{T}} \widehat{\gamma}, \end{aligned} \tag{3}$$

where $\widehat{\gamma}$ is the regression estimator when y is regressed on X_C with weight matrix Ω. Other distance measures have been suggested, but are equivalent asymptotically (Deville and Särndal, 1992) and in practice (Deville *et al.*, 1993).

In practice survey data sets are rarely complete. Calibration may be used to allow for unit non-response, but cases where the covariate x is known but the target variable y is missing demands a different approach, typically through the use of an imputation model that allows missing values of y to be predicted from those available.

A common deterministic approach to imputation is to use a (generalized) linear model based on the vectors x_{hi} of auxiliary variables. The normal equations for estimating the parameters β of such imputation models across strata may be written in the vector form

$$\sum_{h=1}^{H} \sum_{i=1}^{n_h} x_{hi} \psi(y_{hi}, x_{hi}; \beta) = 0, \tag{4}$$

where ψ is the derivative of the implied loss function with respect to β. If the response y is dichotomous it is natural to use logistic regression as the imputation model, and then the y_{hi} are binary indicator variables and $\psi(y, x; \beta) = y - \exp(x^{\mathrm{T}}\beta)/\{1 + \exp(x^{\mathrm{T}}\beta)\}$. If y is continuous then one simple possibility is ratio imputation using a scalar x, for which we take $\psi(y, x; \beta) = y - \beta x$. For a more robust imputation model, one might use Huber's Proposal 2 (Huber, 1981), for which $\psi(u) = \mathrm{sign}(u) \min(|u|, \tau)$; here $\tau > 0$ controls the degree of robustness of the fit, with $\tau \to \infty$ recovering the least squares estimator, and $\tau \to 0$ giving higher robustness. Once the linear model M-estimate $\widehat{\beta}$ of β has been found, the missing response for an individual with explanatory variable x can be predicted by $x^{\mathrm{T}}\widehat{\beta}$, or by a smooth function of this.

For a linear imputation model, the calibrated and imputed Horvitz–Thompson estimator may be written as

$$\begin{aligned}\widehat{\tau} \;=\; & w^{\mathrm{T}}\{Zy + (I-Z)\widehat{y}\} = \omega^{\mathrm{T}}Zy + (c - X_C^{\mathrm{T}}\omega)^{\mathrm{T}}(X_C^{\mathrm{T}}\Omega X_C)^{-1}X_C^{\mathrm{T}}\Omega Zy \\ & + \omega^{\mathrm{T}}(I-Z)X\widehat{\beta} + (c - X_C^{\mathrm{T}}\omega)^{\mathrm{T}}(X_C^{\mathrm{T}}\Omega X_C)^{-1}X_C^{\mathrm{T}}\Omega(I-Z)X\widehat{\beta}, \qquad (5)\end{aligned}$$

where $Z = \mathrm{diag}(z)$ is the $n \times n$ diagonal matrix of indicator variables z_{hi} corresponding to observed response, X is the $n \times q$ matrix that contains the auxiliary variables corresponding to both respondents and nonrespondents, and $\widehat{y} = X\widehat{\beta}$ represents the $n \times 1$ vector of fitted values from the regression model used for imputation.

3 Resampling variance estimation

Modern sample survey estimators often involve calibration and/or imputation, and variance formulae for them cannot be found in classic books such as Cochran (1977). The simplest approach would be to treat the imputed responses $\widehat{y}$ as if they were true responses, but this can lead to considerable underestimation of the true variance. One way to estimate the variance of estimators such as (5) is to turn to resampling. The adaptation of resampling methods to the survey setting requires special care, because it must take into account the complex dependence structures induced by the probability sampling scheme as well as any calibration and imputation procedures. We now briefly outline the main resampling procedures proposed for sample surveys.

The jackknife, originally introduced as a method of bias estimation (Quenouille, 1949a,b) and subsequently proposed for variance estimation (Tukey, 1958), involves the systematic deletion of groups of units at a time, the recomputation of the statistic with each group deleted in turn, and then the combination of all these recalculated statistics. The simplest jackknife entails the deletion of single observations, but this delete-one jackknife is inconsistent for non-smooth estimators, such as the median and other estimators based on quantiles (Efron, 1982). Shao and Wu (1989) and Shao and Tu (1995) have shown that the inconsistency can be repaired by deleting groups of d observations, where $d \to \infty$ as $n \to \infty$. Rao and Shao (1992) describe a consistent version of the delete-one jackknife variance estimator using a particular hot deck imputation mechanism to account for non-response; see also Fay (1996) for a wider perspective, and Chen and Shao (2001), who show that this approach fails for another popular imputation scheme, nearest-neighbour imputation.

The bootstrap involves recomputing the statistic, now using resampling from an estimated population $\widehat{F}$ to obtain bootstrap samples that may be represented by $\widehat{F}^*$, giving corresponding statistics $\widehat{\theta}^* = t(\widehat{F}^*)$. Repeating this process R times independently yields bootstrap replicates $\widehat{\theta}_1^*, \ldots, \widehat{\theta}_R^*$ of $\widehat{\theta}$, and the bootstrap estimate of variance is given by $(R-1)^{-1}\sum_r(\widehat{\theta}^*_r - \overline{\widehat{\theta}^*})^2$. For stratified data, the resampling is performed independently within each stratum. The standard bootstrap uses sampling with replacement, corresponding to independent sampling from an original population, but this does not match the without-replacement sampling generally used in the survey context, so the finite sampling correction is missed, leading to a biased variance estimator. This failure of the conventional bootstrap has spurred a good deal of work on modified bootstraps including the without-replacement bootstrap (Gross, 1980; Chao and Lo, 1985; Bickel and Freedman, 1984; Sitter, 1992b; Booth *et al.*, 1994; Booth and Hall, 1994), the with-replacement bootstrap (McCarthy and Snowden, 1985), the rescaling bootstrap (Rao and Wu, 1988), and the mirror-match bootstrap (Sitter, 1992a). When responses are missing, the imputation mechanism must be applied to each resample $\widehat{F}^*$ (Shao and Sitter, 1996). Thus we must re-impute repeatedly using the respondents of the bootstrapped sample to fit the imputation model and then impute the nonrespondents of the bootstrap sample. This is computer intensive, but it gives consistent variance estimators for medians and other estimators based on quantiles.

Balanced half-sampling (McCarthy, 1969) is the simplest form of balanced repeated replication. It was originally developed for stratified multistage designs with two primary sampling units drawn with replacement in the first stage. Two main generalizations to surveys with more than $n_h = 2$ observations per stratum have been proposed. The first, investigated by Gurney and Jewett (1975), Gupta and Nigam (1987), Wu (1991) and Sitter (1993), uses orthogonal arrays but requires a large number of replicates, making it impractical for many applications. The second generalization, a simpler more pragmatic approach, is to group the primary sampling units in each stratum into two groups, and to apply balanced repeated replication using the groups rather than individual units (Rao and Shao, 1996; Wolter, 1985, Section 3.7). The balanced repeated replication variance estimator v_{BRR} can be highly variable. A solution to this suggested by Robert Fay of the US Bureau of the Census (Dippo *et al.*, 1984; Fay, 1989) is to use a milder reweighting scheme. Another solution (Rao and Shao, 1996) is to repeat the method over differently randomly selected groups to provide several estimates of variance, averaging of which will provide a more stable overall variance estimate. Shao *et al.* (1998) adjust balanced repeated replication to the presence of nonresponse, by taking into account a deterministic or random imputation mechanism. Under a general stratified multistage sampling design, they establish consistency of the adjusted balanced repeated replication variance estimators for functions of smooth and nonsmooth statistics.

Multiple imputation (Rubin, 1987) has also been promoted for variance estimation in complex surveys—standard formulae are computed for several datasets for which missing data have been stochastically imputed, and are then combined in such a way as to make proper allowance for the effect of imputation. This approach has been regarded as controversial by certain authors; see for example Fay (1996).

4 Jackknife linearization

A general approach to construction of linearization variance estimators is though the influence function (Hampel *et al.*, 1986). In many cases the estimand θ can be written as a functional $t(F)$ of the underlying distribution function F. A simple estimator of $t(F)$ is then $t(\widehat{F})$, where $\widehat{F}$ is the empirical distribution function of the data. For the mean, for instance, $t(F) = \int y dF(y)$ and $t(\widehat{F}) = \overline{Y}$ is its empirical analogue. In the case of simple random sampling with replacement, and assuming some differentiability properties for $t(\cdot)$, the estimate $\widehat{\theta} = t(\widehat{F})$ can be expanded around $\theta = t(F)$ as $t(\widehat{F}) \doteq t(F) + n^{-1} \sum_{i=1}^{n} L_t(Y_i; F)$, where

$$L_t(y; F) = \lim_{\epsilon \to 0} \frac{t\{(1-\epsilon)F + \epsilon\delta_y\} - t(F)}{\epsilon}$$

is the *influence function* for $t(\widehat{F})$, δ_y being the distribution function putting a point mass at y. This expansion can be used to establish that the estimator is asymptotically unbiased and Gaussian. Its variance $v_L(F) = n^{-1}\mathrm{var}\{L_t(Y; F)\}$ can be estimated by

$$\widehat{v}_L = n^{-2} \sum_{i=1}^{n} l_i^2, \tag{6}$$

where $l_i = L_t(y_i; \widehat{F})$ are the *empirical influence values* for the statistical functional t evaluated at y_i and $\widehat{F}$. Here l_i can be thought of as the derivative of t at $\widehat{F}$ in the direction of a distribution putting more mass on the i-th observation.

For stratified random sampling without replacement (6) may be modified to

$$v_L = \sum_{h=1}^{H} (1 - f_h) \frac{1}{(n_h - 1)n_h} \sum_{i=1}^{n_h} l_{hi}^2, \tag{7}$$

where l_{hi} is the empirical influence value corresponding to the ith observation in stratum h.

We now consider the Horvitz–Thompson estimator and give formulae for its empirical influence functions for stratified sampling in three situations of increasing complexity:

- the standard estimator (1), for which

$$l_{hi} = n_h \omega_{hi} y_{hi} - \omega_h^{\mathrm{T}} y_h;$$

- the calibrated estimator (3), for which (Canty and Davison, 1999)

$$\begin{aligned} l_{hi} \;\; = \;\; & (n_h \omega_{hi} y_{hi} - \omega_h^{\mathrm{T}} y_h) + (X_{Ch}^{\mathrm{T}} \omega_h - n_h \omega_{hi} x_{Chi})^{\mathrm{T}} \widehat{\gamma} \\ & + n_h \omega_{hi} (c - X_C^{\mathrm{T}} \omega)^{\mathrm{T}} (X_C^{\mathrm{T}} \Omega X_C)^{-1} x_{Chi} (y_{hi} - x_{Chi}^{\mathrm{T}} \widehat{\gamma}), \end{aligned}$$

  where ω_h and y_h are $n_h \times 1$ vectors of the weights and responses for the h-th stratum, X_{Ch} is the $n_h \times q_C$ matrix of calibration covariates for the h-th stratum, and $\widehat{\gamma} = (X_C^{\mathrm{T}} \Omega X_C)^{-1} X_C^{\mathrm{T}} \Omega y$; and

- the calibrated estimator (5) with imputation of missing responses. Let

$$\widehat{\gamma}_M = (X_C^{\mathrm{T}}\Omega X_C)^{-1}X_C^{\mathrm{T}}\Omega(I-Z)\widehat{y}$$

correspond to $\widehat{\gamma}$, but for those individuals with missing responses, and let $l_i(\widehat{\beta})$ be the elements of the $q \times 1$ vector of influence functions for the imputation regression coefficients, corresponding to differentiation with respect to the ith case in stratum h. Then calculations along the lines of those in Rust and Rao (1996) or Canty and Davison (1999) yield

$$\begin{aligned} l_{hi} \;=\; & (n_h\omega_{hi}z_{hi}y_{hi} - \omega_h^{\mathrm{T}}Z_hy_h) + (X_{Ch}^{\mathrm{T}}\omega_h - n_h\omega_{hi}x_{Chi})^{\mathrm{T}}\widehat{\gamma} \\ & + n_h\omega_{hi}(c - X_C^{\mathrm{T}}\omega)^{\mathrm{T}}(X_C^{\mathrm{T}}\Omega X_C)^{-1}x_{Chi}(z_{hi}y_{hi} - x_{Chi}^{\mathrm{T}}\widehat{\gamma}) \\ & + \{n_h\omega_{hi}(1-z_{hi})\widehat{y}_{hi} - \omega_h^{\mathrm{T}}(I_h - Z_h)\widehat{y}_h\} + \omega^{\mathrm{T}}(I-Z)Xl_i(\widehat{\beta}) \\ & + (X_{Ch}^{\mathrm{T}}\omega_h - n_h\omega_{hi}x_{Chi})^{\mathrm{T}}\widehat{\gamma}_M \\ & + n_h\omega_{hi}(c - X_C^{\mathrm{T}}\omega)^{\mathrm{T}}(X_C^{\mathrm{T}}\Omega X_C)^{-1}x_{Chi}\{(1-z_{hi})\tilde{y}_{hi} - x_{Chi}^{\mathrm{T}}\widehat{\gamma}_M\} \\ & + (c - X_C^{\mathrm{T}}\omega)^{\mathrm{T}}(X_C^{\mathrm{T}}\Omega X_C)^{-1}X_{Ch}^{\mathrm{T}}\Omega_h(I_h - Z_h)X_hl_i(\widehat{\beta}). \end{aligned} \quad (8)$$

In particular, use of a linear model fitted by least squares for deterministic imputation yields

$$l_i(\widehat{\beta}) = n_hz_i(X^{\mathrm{T}}ZX)^{-1}x_i(y_i - x_i^{\mathrm{T}}(X^{\mathrm{T}}ZX)^{-1}X^{\mathrm{T}}Zy), \quad i = 1,\ldots,\sum_{h=1}^{H}n_h,$$

where X is the regression matrix. When the regression coefficients vary among the strata, then the $l_i(\widehat{\beta})$ in (8) are taken to be

$$l_i(\widehat{\beta}_h) = n_hz_i(X_h^{\mathrm{T}}Z_hX_h)^{-1}x_i(y_i - x_i^{\mathrm{T}}(X_h^{\mathrm{T}}Z_hX_h)^{-1}X_h^{\mathrm{T}}Z_hy_h),$$

where X_h, Z_h, and y_h are the covariate matrix, the indicator matrix for observed responses, and the response vector for stratum h.

The advantages of such formulae over resampling techniques are a reduction in computational effort and the possibility of handling massive surveys. In the next section we describe the results of a Monte Carlo simulation to compare these procedures.

5 Numerical comparison

Using a realistic simulation based on the 1998 Swiss Household Budget Survey (Renfer, 2001), we consider the calibrated and imputed Horvitz–Thompson estimator of the total expenditure on bread and cereal products, based on complete data from $N = 9275$ households in $H = 7$ strata of various sizes. Also available on each household is a set of 14 auxiliary variables, of which 10 population margins are known. For the simulation, we consider the $N = 9275$ households as the whole population, for which we assume we know the total expenditure. We perform stratified random sampling without replacement and with equal inclusion probabilities of 1/8 within 6 strata, and 3/8 in the other stratum, giving a sample size of 1332. Item non-response for the response variable is applied using a uniform probability of missingness across the entire sample. On each of the 500 samples simulated, we calculate the calibrated and imputed Horvitz–Thompson estimates, and use various resampling techniques to obtain variances for them.

The bootstrap used 100 replicates of the calibrated and imputed Horwitz–Thompson estimator, obtained by the procedure of Shao and Sitter (1996), that is, with missing responses imputed deterministically using a linear model fitted to the bootstrapped full respondents, and with the imputed dataset calibrated to the weights by linear regression. A separate simulation showed that 100 bootstrap replicates was adequate. To match the computational complexity of the bootstrap, we used roughly the same number of block deletions when applying the block jackknife with replacement. This was applied with 13 randomly-selected blocks in each stratum, leading to about 91 computations in all for each jackknife variance estimate. Two forms of balanced repeated replication were applied, the first using a single random split of each stratum into two halves for each replication; no Fay factor was used but the weights for those observations included in the replicate were multiplied by a factor of two before calibration. The second form, repeatedly-grouped balanced repeated replication, averages over variance estimates from 13 such splits. The jackknife linearization estimator is that given by (7) and (8).

The standard formulae for multiple imputation were applied, using 30 random imputations from a linear model fitted to the complete data; for parametric imputation we used a homoscedastic normal error model, with the values of the regression parameters and variance changing randomly and independently according to the fitted normal and chi-squared distributions between simulations; for nonparametric imputation errors were simulated according to a model-based residual bootstrap (Davison and Hinkley, 1997, page 262).

Table 1 and Figures 1 and 2 compare the performances of these variance estimation techniques for missingness rates of 0%, 20%, 40%, and 60%. Linearisation and both multiple imputation methods give the same results when no data are missing. The block jackknife underestimates the true variances, which are systematically overestimated by repeatedly-grouped balanced repeated replication. Balanced repeated replication without grouping is highly variable by comparison, in agreement with results of Rao and Shao (1996), but grouped balanced repeated replication works much better. Jackknife linearisation works well for low levels of missingness, and overall it produces variances that are rather too low but quite stable. The bootstrap performs best of all for higher levels of missingness. Nonparametric multiple imputation also performs well.

Figure 2 shows how the variance estimates for the 500 simulated data sets are correlated with the bootstrap variance estimates. Linearization, repeatedly grouped balanced repeated replication and multiple imputation variance estimates are fairly closely correlated with bootstrap variance estimates. The added variability of the variances from balanced repeated replication shows clearly.

Proportion missing (%)	Relative bias (%) 0	20	40	60	Relative RMSE (%) 0	20	40	60
Block jackknife	−10	−11	−16	−15	1820	2210	3340	4120
Balanced repeated replication (BRR)	3	5	−3	3	3830	4660	4900	6510
Randomly grouped BRR	8	7	2	6	1710	1830	1980	2960
Bootstrap	7	5	−1	0	1510	1580	1840	2480
Linearization	−0.3	−3	−9	−10	800	1040	2050	2750
Multiple imputation, parametric	−0.3	0.1	−7	−14	800	1040	1900	3500
Multiple imputation, nonparametric	−0.3	11	11	8	800	1950	2490	2560

Table 1: Relative bias and root mean squared error (%) for the different resampling plans applied to simulated data based on the 1998 Swiss Household Budget Survey, for different proportions of missing data.

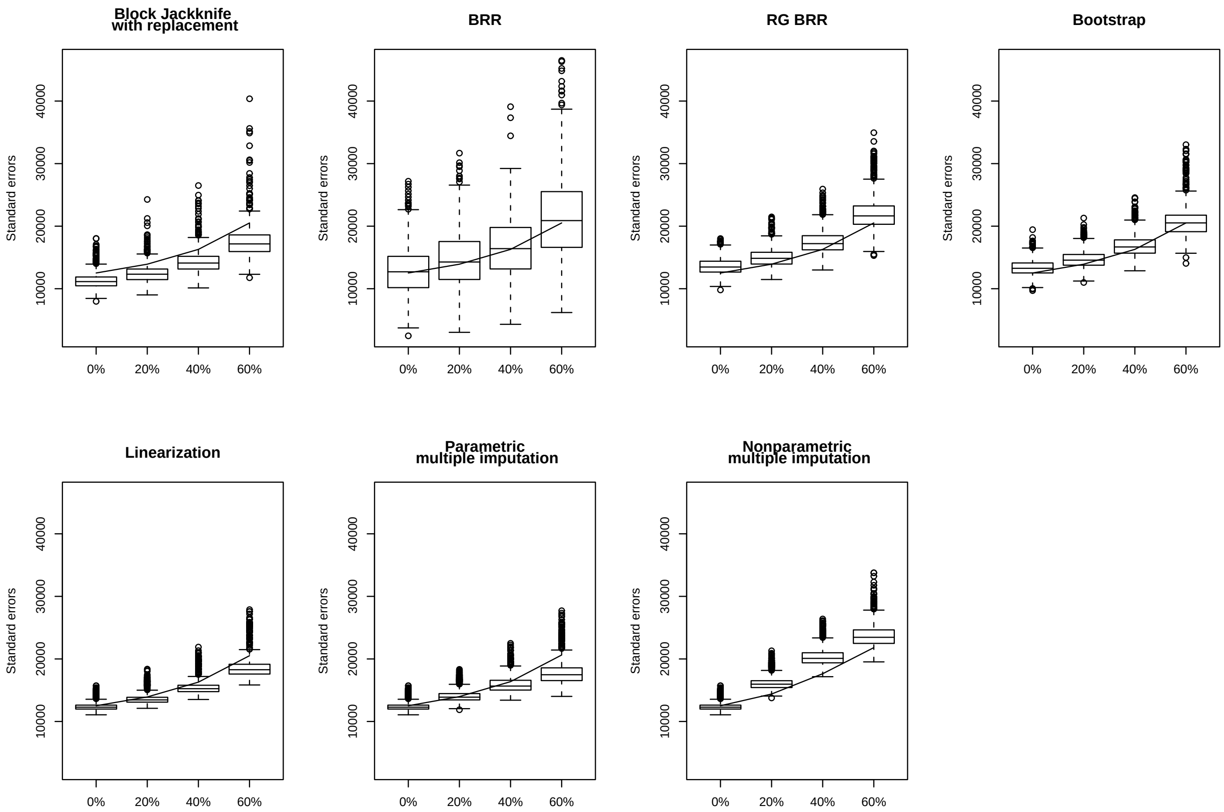

Figure 1: Comparison of resampling estimators of variance in the presence of calibration and imputation, as a function of the proportion of missing data. Simulation based on the 1998 Swiss Household Budget Survey. The solid line shows the true variances, estimated from 10,000 simulations, and the boxplots show the variance estimates computed for 500 samples. RG and BRR indicate repeatedly grouped and balanced repeated replication respectively.

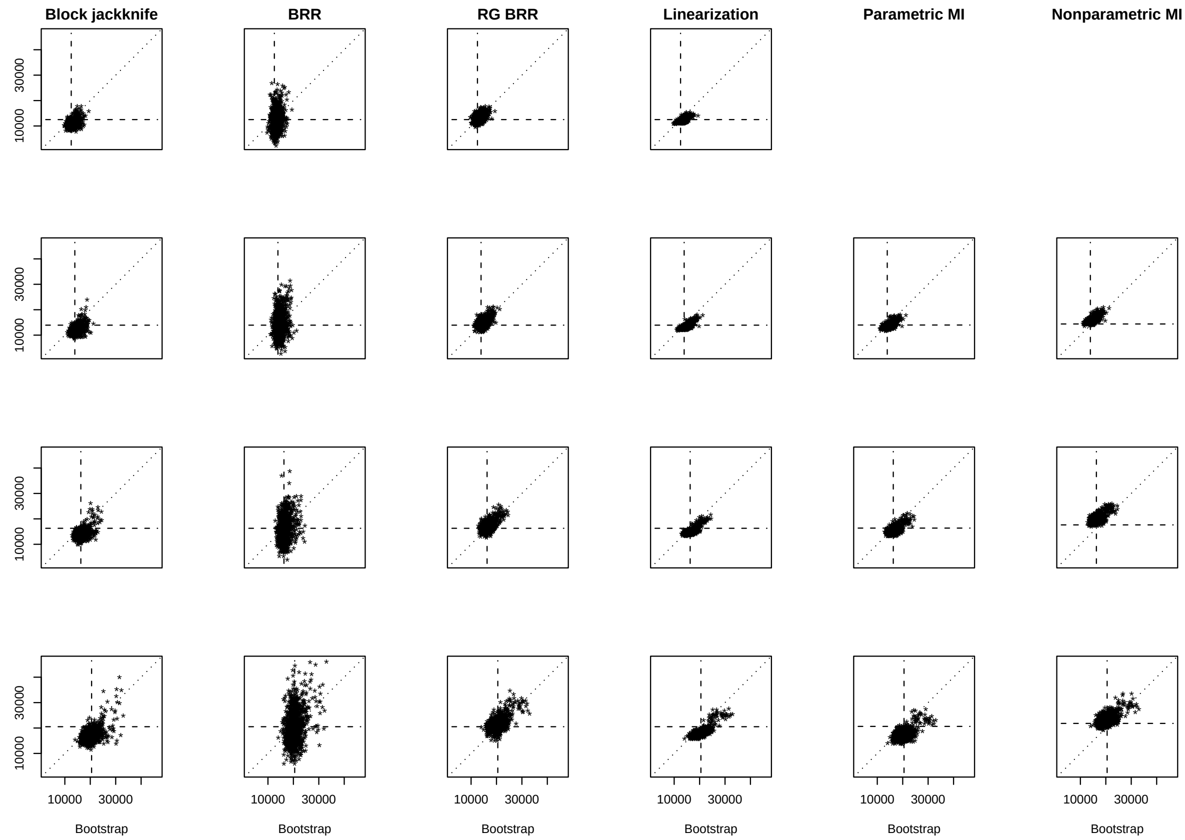

Figure 2: Comparison of resampling standard errors in the presence of calibration and imputation, as a function of the proportion of missing data; from top to bottom 0%, 20%, 40%, 60% item non-response. The dashed lines are the 'true' sampling standard errors, and the dotted line shows $x = y$. Simulation based on the 1998 Swiss Household Budget Survey. RG, BRR and MI indicate repeatedly grouped, balanced repeated replication and multiple imputation respectively.

Overall the bootstrap approach of Shao and Sitter (1996), the linearization method of Section 4, and nonparametric multiple imputation seem best in terms of bias and stability. As far as computation time is concerned, the advantage goes to linearization, which is up to fifty times faster than the other methods included in the study.

6 Discussion

The broad conclusions of the numerical study above support those of Canty and Davison (1999), who concluded that jackknife linearisation and the bootstrap were the simplest and most accurate methods of variance estimation in their study. They did not consider imputation, but found similar conclusions for a variety of smooth estimators and for differences of them between two sampling occasions. It seems reasonable to suppose that the same general results seen above would also extend to a broader context.

References

Berger, Y. G. and Skinner, C. J. (2003) Variance estimation for a low income proportion. *Applied Statistics* **52**, 457–468.

Bickel, P. J. and Freedman, D. A. (1984) Asymptotic normality and the bootstrap in stratified sampling. *Annals of Statistics* **12**, 470–482.

Booth, J. G., Butler, R. W. and Hall, P. (1994) Bootstrap methods for finite populations. *Journal of the American Statistical Association* **89**, 1282–1289.

Booth, J. G. and Hall, P. (1994) Monte Carlo approximation and the iterated bootstrap. *Biometrika* **81**, 331–340.

Canty, A. J. and Davison, A. C. (1999) Resampling-based variance estimation for labour force surveys. *The Statistician* **48**, 379–391.

Chao, M. T. and Lo, S. H. (1985) A bootstrap method for finite populations. *Sankyā A* **47**, 399–405.

Chen, J. and Shao, J. (2001) Jackknife variance estimation for nearest-neighbor imputation. *Journal of the American Statistical Association* **96**, 260–269.

Cochran, W. G. (1977) *Sampling Techniques*. Third edition. New York: Wiley.

Davison, A. C. and Hinkley, D. V. (1997) *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

Deville, J. C. and Särndal, C. E. (1992) Calibration estimators in survey sampling. *Journal of the American Statistical Association* **87**, 376–382.

Deville, J. C., Särndal, C. E. and Sautory, O. (1993) Generalized raking procedures in survey sampling. *Journal of the American Statistical Association* **88**, 1013–1020.

Dippo, C. S., Fay, R. E. and Morganstein, D. H. (1984) Computing variances from complex samples with replicate weights. In *Proceedings of the Section on Survey Research Methods*, pp. 489–494. Washington DC: American Statistical Association.

Efron, B. (1982) *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadephia: SIAM.

Fay, R. E. (1989) Theory and application of replicate weighting for variance calculations. In *Proceedings of the Social Statistics Section*, pp. 212–217. American Statistical Association.

Fay, R. E. (1996) Alternative paradigms for the analysis of imputed survey data. *Journal of the American Statistical Association* **91**, 490–498.

Gross, S. (1980) Median estimation in sample surveys. In *Proceedings of the Section on Survey Research Methods*, pp. 181–184. Alexandria, VA: American Statistical Association.

Gupta, V. K. and Nigam, A. K. (1987) Mixed orthogonal arrays for variance estimation with unequal numbers of primary selections per stratum. *Biometrika* **74**, 735–742.

Gurney, M. and Jewett, R. S. (1975) Constructing orthogonal replications for standard errors. *Journal of the American Statistical Association* **70**, 819–821.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986) *Robust Statistics: The Approach Based on Influence Functions.* New York: Wiley.

Huber, P. J. (1981) *Robust Statistics.* New York: Wiley.

McCarthy, P. J. (1969) Pseudo-replication: Half samples. *Review of the Interational Statistics Institute* **37**, 239–264.

McCarthy, P. J. and Snowden, C. B. (1985) The bootstrap and finite population sampling. *Vital and Health Statistics* **2**, 2–95.

Quenouille, M. H. (1949a) Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society, Series B* **11**, 68–84.

Quenouille, M. H. (1949b) Notes on bias in estimation. *Biometrika* **43**, 353–360.

Rao, J. N. K. and Shao, J. (1992) Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* **79**, 811–822.

Rao, J. N. K. and Shao, J. (1996) On balanced half-sample variance estimation in stratified random sampling. *Journal of the American Statistical Association* **91**, 343–348.

Rao, J. N. K. and Wu, C. F. J. (1988) Resampling inference with complex survey data. *Journal of the American Statistical Association* **83**, 231–241.

Renfer, J.-P. (2001) *Description and process of the Household and Budget Survey of 1998 (HBS 1998).* Swiss Federal Statistical Office. 1-19.

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

Rust, K. F. and Rao, J. N. K. (1996) Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research* **5**, 283–310.

Shao, J., Chen, Y. and Chen, Y. (1998) Balanced repeated replication for stratified multistage survey data under imputation. *Journal of the American Statistical Association* **93**, 819–831.

Shao, J. and Sitter, R. R. (1996) Bootstrap for imputed survey data. *Journal of the American Statistical Association* **91**, 1278–1288.

Shao, J. and Tu, D. (1995) *The Jackknife and Bootstrap.* New York: Springer-Verlag.

Shao, J. and Wu, C. F. J. (1989) A general theory for jackknife variance estimation. *Annals of Statistics* **17**, 1176–1197.

Sitter, R. R. (1992a) A resampling procedure for complex survey data. *Journal of the American Statistical Association* **87**, 755–765.

Sitter, R. R. (1992b) Comparing three bootstrap methods for survey data. *Canadian Journal of Statistics* **20**, 135–154.

Sitter, R. R. (1993) Balanced repeated replications based on orthogonal multi-arrays. *Biometrika* **80**, 211–221.

Tukey, J. W. (1958) Bias and confidence in not quite large samples (abstract). *Annals of Mathematical Statistics* **29**, 614.

Wolter, K. M. (1985) *Introduction to Variance Estimation.* New York: Springer-Verlag.

Wu, C. F. J. (1991) Balanced repeated replications based on mixed orthogonal arrays. *Biometrika* **78**, 181–188.