Methods for variance estimation under random hot deck imputation in business surveys

Paolo Righi Stefano Falorsi Andrea Fasulo * †

Abstract

When imputed values are treated as if they were observed, the precision of the estimates is generally overstated. In the paper three variance methods under imputation are taken into account. Two of them are the wellknown bootstrap and Multiple Imputation. The third is a new method based on grouped jackknife easy to implement, not computer intensive and suitable when random hot deck imputation is performed. A simulative comparison on real business data has been carried out. The findings show that the proposed method has good performances with respect to the other two.

Keywords: Bootstrap, Multiple Imputation, Jacknife, Extended DAGJK, Replicate weights, Monte Carlo simulation

1. Introduction

Variance estimation has to take into account an additional complexity element: the unit and item nonresponse that commonly trouble the large scale surveys. Unit nonresponse is customarily handled by forming weighting classes using auxiliary variables observed on all the sampled elements. Then adjusting the survey weights of all respondents within a weighting class by a common nonresponse adjustment factor, with different adjustment factors in different classes (Kalton and Kasprzyk 1986).

Imputation is the commonly used approach to compensate for missing (item nonresponse) or invalid values in sample surveys (Kalton and Kasprzyk 1986). In the paper the random hot deck imputation is considered.

When unit and item nonresponse correction is performed extra variability is introduced in the sampling errors. Modifications of Taylor and resampling methods for contemplating unit nonresponse are quite straightforward while item nonresponse is a ticklish issue. Analyses performed on imputed values treated as if they were observed, can be misleading when estimates of the variance do not include the variability component due to imputation. As a result, the precision of estimates is overstated, and subsequent statistical analyses can be misleading (e.g., confidence intervals have lower than nominal levels).

The approaches proposed in literature to obtain valid variance estimators in presence of imputed data are divided according to several classifications. A first common classification distinguishes among linearization (or Model Assisted techniques see Särndal 1992), resampling (Shao and Tu 1995; Wolter 2007) and the Multiple Imputation (Rubin 1987) methods being the first two categories used for the complete data variance estimates as well.

The resampling techniques in presence of item nonresponse can be divided according to the standard classification used for the complete data (Wolter 2007). Then, bootstrap (Efron 1994; Shao and Sitter 1996; Saigo et al. 2001; Shao 2003), balanced repeated replication

* Italian National Statistical Insitute (Istat), e-mail: parighi@istat.it; stfalors@istat.it; fasulo@istat.it.

† The views expressed in this paper are solely those of the authors and do not involve the responsibility of Istat.

(Rao and Shao 1999), random group (Shao and Tang 2001) and jackknife (Rao and Shao 1992; Rao 1996; Yung and Rao 2000; Chen and Shao 2001; Skinner and Rao 2002; Saigo 2005) methods, can be distinguished.

In the literature there is not a common judgement on which is the best approach or method. This is the main reason why all these methods are investigated in a lot of different contexts involving the sampling design, the estimator, the domains of interest and the imputation process. Furthermore the dimension of the survey and the type and the number of parameters to be estimated have to be taken into account. For large scale surveys, the function to be estimated, the complexity of sampling design, the imputation procedure and the cost-effectiveness issues drive the choice of a specific variance estimator.

The paper focuses on the operational conditions for implementing the methods in the data production in Official Statistics. In particular two issue are raised: the setting up of theoretical framework for applying the method; the computational aspect, that is always troublesome for large scale surveys.

Three methods are deeply investigated: the bootstrap under imputation, the Multiple Imputation and finally a new method based on a jakknife technique. The linearization is kept out because can be problematic to consider in a standardized data production process in which the timeliness is a pressing data quality dimension.

Section 2. introduces the three methods. In particular, section 2.1 gives a literature review on the jackknife techniques with the random hot deck imputation. Section 2.2 proposes a new variance estimator taking into account the item non responses based on a grouped jackknife technique. This is the innovative output of the paper. Section 2.3 and 2.4 are respectively devoted to the bootstrap estimator and Multiple Imputation procedure. In section 3. we compare the three methods by means of a Monte Carlo simulation based on real business survey data. Some concluding remarks are given section 4.

2. Bootstrap, Multiple Imputation and jacknife techniques with random hot deck imputation

The variance estimation process under imputation depends on the kind of imputation procedure has been used. For instance, if a jackknife type estimator has been chosen, the form of the final estimator can change according to the imputation.

In the paper the hot deck procedure is taken into account. This class of methods is one of the most popular in the survey sampling. Various specifications of the method are proposed in the literature. In the simplest form of hot deck imputation, a random sample of size $\bar{r}$ (the number of nonrespondents) is selected from the sample of respondents to an item y, and the associated item y values are used as donors. The accuracy of imputation depends on the nonresponse model (the imputation classes) and on the simple or weighted random selection of the donors.

Since the hot deck imputation is a form of regression imputation (Kalton and Kasprzyk 1986) the analysis of the variance estimators with this imputation technique is not so restrictive.

Eventually, we consider the standard imputation procedures based on real observed values. Hence, we do not explore variance estimators for the imputation procedures where imputed values depend on variables already imputed in previous steps (Raghunathan et al. 2001). That means the hot deck imputation classes are defined on the observed values.

2.1 Jackknife variance methods under imputation

Jackknife variance estimation in presence of item non response has been extensively studied in literature.

One of the first papers was made by Burns (1990). He proposed to perform a new imputation for each jackknife sample according to the same procedure applied on the overall sample. Then, being n the sample size, the procedure needs $n + 1$ imputation steps.

Rao and Shao (1992) showed that the procedure can lead to serious overestimation for large sample size. They propose a consistent jackknife variance estimator in presence of imputed data (for the Horvitz-Thompson estimator) by means of hot deck methods assuming equal response probabilities within imputation classes. The method is suitable for stratified random sampling and stratified multistage sampling design even in the general case in which the imputation classes cut across the sampled clusters. The approach, named *adjusted jackknife*, performs only one single imputation on the full sample and it adjusts the imputed values for each pseudo-replicate before applying the standard jackknife variance formula for stratified design. Then, the technique is much more efficient in terms of computation with respect to the Burns approach.

The consistency of the adjusted jackknife (for smooth functions such as totals and means) is shown assuming equal response probabilities within imputation class and performing independently within each class the hot deck imputation.

In presence of variable inclusion probabilities in the stratum (such as in the multistage sampling designs) the properties holds when the weighted hot deck is implemented. Weighted hot deck select a donor from the imputation cell with probability proportional to the sampling weight. Conversely if a simple random sampling is used to select a donor then the estimator, under imputation, of the target parameter will be biased and it's variance estimator as well.

The adjusted jackknife method needs for each unit an imputation flag on the data set.

Further enhancements of the adjusted jackknife are given by Rao and Sitter (1995) that examine the jackknife with ratio imputation in the model based framework. Rao (1996) gives new results about the adjusted jackknife variance estimator with imputed survey data. As far simple random sampling is concerned, Rao shows some properties when ratio and regression imputation are used for estimating totals or means.

In particular, the adjusted jackknife variance estimator is design consistent and it is also design and model unbiased under the imputation model.

Regarding stratified multistage sampling Rao shows the properties of the adjusted jackknife variance estimator when the mean imputation in the imputation classes is used. Within each imputation class, the weighted mean of the interest variable computed on the respondents is assigned to all missing responses. The technique assumes the best predictor of the missing values is obtained by a homoscedastic mean superpopulation model. If the model holds for the respondents and under uniform response within each class the adjusted jackknife is design consistent.

Yung and Rao (2000) extend the analysis of the adjusted jackknife variance estimator under imputation when poststratified or generalized regression estimator are used. The weighted mean imputation and weighted hot deck stochastic imputation within imputation classes have been studied.

When the weighting classes are the poststrata, the estimator and corresponding jackknife variance estimator are simply computed on the respondents. The authors show also the jackknife estimator when weighting classes cut across the poststrata and they give the proof of the asymptotic consistency. The property holds also when generalized regression estimator and weighted hot deck imputation are used.

Furthermore, the authors investigate the properties of the jackknife variance estimators under weighting adjustment for unit non response. The properties are essentially studied in the stratified multistage sampling design. In case of unit non response, the authors assume a set of weighting classes. Within each class a uniform response mechanism is supposed. Moreover, non response adjustment is performed before the poststratification adjustment so that the known totals are benchmarked.

This comprehensive theoretical framework encompassing general point estimators and unit and item non response makes jackknife techniques appealing .

2.2 The modified Extended DAGJK under hot deck imputation

The adjusted jackknife method has a remarkable reduction of computational effort with respect to the imputation procedure but it is still computer intensive, because of the number of replications in the estimation procedure. For reducing the computations a commun strategy is to combine units into variance strata and perform a grouped jackknife instead of a standard delete one-unit jackknife (Rust 1985; 1986; Rust and Rao 1996). The family of techniques delete groups of units rather than one unit at time for reducing calculation effort. The creation of a replicate group can be done within design stratum or, combining design strata into superstratum, taking groups of units within superstratum that cut across design strata.

There is limited theoretical guidance on how the grouping should be done and much is based on heuristic knowledge. A common assumption in the literature is to form equal sized groups. Moreover, to take into account a nonnegligible sampling fraction, it can be useful to form superstrata with design strata having similar sampling fractions (Valliant et al. 2008). Finally, grouped jackknife methods distinguishes themself by the computation of the replicate weights as well, augmenting the possible variance estimators.

Currently there is no empirical evidence showed in literature, suggesting the best grouped jackknife. Then we should underline the importance of having evidence of the empirical properties of these methods in practical applications in the Official Statistics context.

As concern grouped jackknife methods taking into account item nonresponse few have been written in the literature. Brick et al. (2005) show a grouped adjusted jackknife according to Rao and Shao approach in case of Horvitz-Thompson estimator. Di Zio et al. (2008) propose the definition of the the Rao and Shao adjustment and the Delete A Group Jackknife (DAGJK). Miller and Kott (2011) investigate a DAGJK with imputed data with a different approach.

In the following we introduce a new method combining the DAGJK technique with the Rao and Shao adjustment.

Delete A Group Jackknife (Kott 1998; 2001) is a variance estimation technique computationally less intensive than classical jackknife and it can be applied also in case of large scale surveys. DAGJK is within the strategies aiming at reducing the number of jackknife replications, while maintaining adequate precision of variance estimates. It assumes an unique superstratum formed by all the design strata and the replicate groups have units belonging to different design strata. Then, the method does not present implications on the definition of the groups and does not require analysis to form superstrata. This analysis can become cumbersome for large scale and complex business surveys and may affect the timeliness of data production. For this reason variance estimation techniques implemented by a sort of automated process and leading to *good* statistical results may be preferred to better techniques but more complex to be implemented.

Consider the stratified simple random sampling commonly used in the business surveys, where in each stratum h are included N_h units. A sample of $n_h \geq 2$ units is drawn from each stratum independently across strata. Let $d_{hk}(> 0)$ be the basic weight of unit k in stratum h, denoted as hk, the estimator of the parameter of the total θ is $\hat{\theta} = \sum_{hk \in s} d_{hk} y_{hk}$, being y_{hk} the value of the variable of interest. The DAGJK technique divides the overall or *parent* sample, s into Q mutually exclusive *replicate* or *random* groups, hereinafter denoted by $s^1, \ldots, s^q, \ldots, s^Q$. Given the subsample s^q, the sample sizes in the strata are indicated as $n_1^q, \ldots, n_h^q, \ldots, n_L^q$. The complement of each s^q is called the *jackknife replicate group* $s^{(q)} = s - s^q$, being $n_1^{(q)}, \ldots, n_h^{(q)}, \ldots, n_L^{(q)}$ the strata sample sizes of $s^{(q)}$. The variance estimator is based on the following jackknife procedure:

1. units are randomly ordered in each stratum;
2. from this ordering the units are systematically allocated into Q groups;
3. for each unit hk, Q different sampling weights (*replicate sampling weights*) are computed;
4. given the qth set of the replicate weights the qth replicate estimate is $\hat{\theta}^{(q)} = \sum_{hk \in s} d_{hk}^{(q)} y_{hk}$.
   where $d_{hk}^{(q)}$ denotes the qth replicate weight of unit hkth ;
5. the DAGJK variance estimation is given by

$$v(\hat{\theta}) = \frac{Q-1}{Q} \sum_{q=1}^{Q} (\hat{\theta}^{(q)} - \hat{\theta})^2. \tag{1}$$

The standard DAGJK replicate weights are given by

$$d_{hk}^{(q)} = \begin{cases} d_{hk}, & \text{when } k \in h \text{ and no unit of } h \text{ belongs to group } q \\ 0, & \text{when } \ k \in q \\ [n_h/(n_h - n_h^q)]\, d_{hk} \quad , & \text{otherwise.} \end{cases} \tag{2}$$

There is not an optimal value for Q. When the number of random groups has to be chosen it needs to consider that increasing the number of random groups the variability of the variance estimation is restricted but the computational effort is augmented. In general, it is common practice a choice between 15 and 80 (Kott 1998; Rust 1985), considering that when Q is greater than 15 the Student's t distribution is approximated quite good by the normal distribution.

The statistical properties in terms of bias and variability of the variance estimates depends on the values of Q, n_h and in the case of WOR designs on the sampling fraction in each stratum. In the latter case, with large sampling fractions the (1) produces conservative variance estimates. Neverthless, Kott (2001) shows that even if the finite population correction factor is negligible but $n_h < Q$ for some strata the (1) is still an upward biased variance estimator. For instance, if all $n_h > 5$ but $n_h < Q$ the relative bias of (1) with weights (2) for Horvitz-Thompson estimator is at most 20%. The upperbound of bias is given by $[Q/(Q-1)] max_h \{n_h/[n_h - 1]\}$ which is itself bounded by $max_h \{n_h/[n_h - 1]\}$. The relative upward bias is equal to $max_h \{n_h/[n_h - 1]\} - 1 = max_h \{1/[n_h - 1]\}$.

Kott developed a different expression of the replicate weights defining the Extented DAGJK (EDAGJK). For the Horvitz-Thompson estimator the replicate weights of EDAGJK assume the following expression,

$$d_{hk}^{(q)} = \begin{cases} d_{hk}, & \text{when } k \in h \text{ and no units of } h \text{ belongs to group } q \\ d_{hk}[1 - (n_h - 1)Z] \quad , & \text{when } \ k \in q; \\ d_{hk}(1 + Z), & \text{otherwise.} \end{cases} \tag{3}$$

where $Z^2 = Q/[(Q-1) n_h (n_{h-1})]$.

With the Greg estimator the replicate weights are given by $w_{hk}^{(q)} = d_{hk}^{(q)}\gamma_{hk}^{(q)}$ and the replicate qth GREG estimate is $\hat{\theta}_{greg}^{(q)} = \sum_{hk \in s} y_{hk} w_{hk}^{(q)}$. The correction factor $\gamma_{hk}^{(q)}$ may be calculated according different ways. Let consider the following expressions:

$$\gamma_{hk}^{(q)} = 1 + \left(\mathbf{X} - \sum_{hk \in s} \mathbf{x}_{hk} d_{hk}^{(q)} \right) \left(\sum_{hk \in s} \frac{\mathbf{x}_{hk}\mathbf{x}'_{hk} d_{hk}^{(q)}}{c_{hk}} \right)^{-1} \frac{\mathbf{x}_{hk}}{c_{hk}}. \tag{4}$$

and

$$\gamma_{hk}^{(q)} = \gamma_{hk} + \left(\mathbf{X} - \sum_{hk \in s} \mathbf{x}_{hk} d_{hk}^{(q)} \gamma_{hk} \right) \left(\sum_{hk \in s} \frac{\mathbf{x}_{hk}\mathbf{x}'_{hk} d_{hk}^{(q)} \gamma_{hk}}{c_{hk}} \right)^{-1} \frac{\mathbf{x}_{hk}\gamma_{hk}}{c_{hk}}. \tag{5}$$

Kott (2006) offers some suggestions on the factor has to be used.

In order to take into account item nonresponse in variance estimation, we propose a modified version of EDAGJK based on the Rao and Shao adjustment for hot deck imputation.

The modified variance estimator is :

$$v(\hat{\theta}_I) = \frac{Q-1}{Q} \sum_{q=1}^{Q} (\hat{\theta}_I^{(q)} - \hat{\theta}_I)^2 \tag{6}$$

where,

$$\hat{\theta}_I = \sum_{hk \in s_R} w_{hk} y_{hk} + \sum_{hk \in s_{\bar{R}}} w_{hk} y_{hk}^*, \tag{7}$$

is the estimator with imputed hot deck values y_{hk}^*, being s_R and $s_{\bar{R}}$ the sample of respondents and non respondents.

$\hat{\theta}_I^{(q)}$ is defined as

$$\hat{\theta}_I^{(q)} = \sum_{g=1}^{G} \left\{ \sum_{hk \in s_{R_g}} w_{hk}^{(q)} y_{hk} + \sum_{hj \in s_{\bar{R}_g}} w_{hj}^{(q)} \left(y_{hj}^* + \hat{y}_{Rg}^{(q)} - \bar{y}_{Rg} \right) \right\} \tag{8}$$

in which: g $(g = 1, \ldots, G)$ indicates the gth imputation cell; s_{R_g} and $s_{\bar{R}_g}$ are respectively the respondents and non repondents in the cell g; $w_{hk}^{(q)}$ are the replicate base or Greg weights. Finally, $\hat{y}_{Rg}^{(q)} = \sum_{hj \in s_{R_g}} w_{hj}^{(q)} y_{hj} / \sum_{hj \in s_{R_g}} w_{hj}^{(q)}$ and $\bar{y}_{Rg}^{(q)} = \sum_{hj \in s_{R_g}} w_{hj} y_{hj} / \sum_{hj \in s_{R_g}} w_{hj}$.

Note that the imputation procedure is performed only on the parent sample.

2.3 Bootstrap variance methods under imputation

Bootstrap method in presence of imputed data has been deeply studied in the relevant papers by Efron (1994) and Shao and Sitter (1996).

Starting from the evidence that the naive approach (treating the imputed values as observed values and using the standard bootstrap) does not capture the inflation in variance due to imputation and serious variance underestimation is possible they showed some procedures together with reimputing bootstrap datasets. In particular circumstances such approaches define a valid approximation to the distribution of $\hat{\theta}_I$. Let Y be the observed data set, being the estimator $\hat{\theta} = f(Y)$, and let $\hat{\theta}_I$ the estimator with imputed values given by the (7), the bootstrap variance estimator replaces B bootstrap estimates, $\hat{\theta}^{(b)}$ $(b = 1, \ldots, B)$, by $\hat{\theta}_I^{(b)}$, where the index (b) denote the estimate based on the bth resampling. Each estimate $\hat{\theta}_I^{(b)}$ is computed according to the same procedure implemented for the overall sample.

For the bth bootstrap sample the procedure by Shao and Sitter can be described as follows:

1. draw a simple random sample $\left\{y_{hk}^{(b)} : k = 1, \ldots, n_h - 1\right\}$ (where y_{hk} denotes the value of y for the unit k beloging to stratum h) with replacement from the sample $\{\tilde{y}_{hk} : k = 1, ..., n_h\}$, independently across the strata, where $\tilde{y}_{hk} = \{y_{hk} : (hk)\} \in s_R \cup \{y_{hk}^* : (hk) \in s_{\bar{R}}\}$
2. apply the same imputation procedure used in costructing the imputed survey data. Denote the bootstrap analogue of $\hat{\theta}_I$ by $\hat{\theta}_I^{(b)}$

$$\hat{\theta}_I^{(b)} = \sum_{s_R^{(b)}} w_k^{(b)} y_k + \sum_{s_{\bar{R}}^{(b)}} w_k^{(b)} y_k^{*(b)}. \quad (9)$$

   where $y_k^{*(b)}$ is the imputed value using the bth bootstrap data and $w_k^{(b)}$ is $n_h/(n_h - 1)$ times the survey weight of unit k.

The bootstrap variance estimator $v(\hat{\theta}_I)$ when has no explicit form may be approximated by

$$v(\hat{\theta}_I) \approx \frac{1}{B} \sum_{b=1}^{B} \left(\hat{\theta}_I^{(b)} - \bar{\hat{\theta}}_I^{(b)}\right)^2 \quad (10)$$

in which $\bar{\hat{\theta}}_I^{(b)} = (1/B) \sum_{b=1}^{B} \hat{\theta}_I^{(b)}$ with $b = 1, \ldots, B$.

Efron (1994) shows that the process generates asymptotically valid variance and distribution estimator for complex sampling designs. The same result was established in Shao and Sitter (1996) for stratified sampling with large n_h. The assumption of negligible sampling fraction in each stratum means that the procedure give consistent variance estimator when with replacement sampling design is implemented. Nevertheless, in business surveys is not uncommon to define very detailed strata with small sample size. In such cases some precautions must be taken. The problems is known also when complete data set is used and a stratified random sampling without replacement is the utilized design. A simple and heuristic approach is to collapse the original strata forming variance strata. Shao and Sitter investigate some bootstrap methods that deal with the small n_h's. In particular they analyze the rescaling bootstrap proposed by Rao and Wu (1988) showing that the method does not work with imputed values.

The simulation study carried out by the authors produces valid approximation with deterministic imputation. In case of random imputation, such as hot deck imputation, upward biased estimates are obtained when some n_h are very small.

To overcome the problem Saigo et al. (2001) have developed a third type of modified bootstrap, the repeated half-sample bootstrap, which together with reimputing bootstrap data sets produces a valid approximation of the distribution of $\hat{\theta}_I$, regardless of whether the imputation is random or not and whether n_h is small or not. In the paper by Shao (2003) are well illustrated the different bootstrap methods and their associated problems.

Eventually, in Shao and Sitter (1996) for avoiding the complete reimputation process for each replication has been proposed a slightly different bootstrap method. Such technique preserves the asymptotic properties except for variance estimate of quantiles.

2.4 Multiple Imputation

Multiple Imputation (MI) was first proposed and thoroughly described in Rubin (1978). More recently the book by Little and Rubin (2002) offers a concise and complete description of the method.

MI is a procedure replacing each missing value by an ordered vector composed of $M \geq 2$ possible values. The ordering assumes that the first components of the vectors for the missing values are used to create one completed data set, the second components of the vectors are used to create the second completed data set and so on. Each completed data set is investigated using standard complete-data methods. To analyze the repetitions within one imputation model to yield a valid inference under the posited reasons for missing data, the M complete-data based on the M repeated imputations are then combined to create one repeated-imputation inference.

Let $\hat{\theta}_m$ and W_m $(m = 1, \ldots, M)$ be the mth complete-data estimate and its variance of the parameter θ obtained by imputation under one model for nonresponse. The MI estimate is given by

$$\bar{\theta}_M = \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}_m. \tag{11}$$

The variability associated with this estimate has two components: the average within imputation variance

$$\bar{W}_M = \frac{1}{M} \sum_{m=1}^{M} W_m \tag{12}$$

and the between-imputation component,

$$B_M = \frac{1}{M-1} \sum_{m=1}^{M} \left(\hat{\theta}_m - \bar{\theta}_M \right)^2 \tag{13}$$

where with vector θ the $(\cdot)^2$ is replaced by $(\cdot)^T(\cdot)$.

The total variability associated to $\bar{\theta}_M$ is given by

$$T_M = \bar{W}_M + \frac{M+1}{M} B_M. \tag{14}$$

With scalar parameter the approximate reference distribution for interval estimates and significance tests is a t distribution

$$\left(\theta - \bar{\theta}_M \right) T_M^{-1/2} \sim t_d, \tag{15}$$

where the degrees of freedom,

$$d = (M-1) \left(1 + \frac{1}{M+1} \frac{\bar{W}_M}{B_M} \right)^2. \tag{16}$$

are based on the Satterthwaite approximation (Rubin and Schenker 1986) . An improved expression of the degree of freedom for small data sets is given, for example, in Little and Rubin (2002).

An estimate of the fraction of missing information γ_M about θ due to nonresponse is given by

$$\hat{\gamma}_M = (1 + 1/M)\frac{B_M}{T_M}. \tag{17}$$

Rubin (1987), shows that the relative efficiency of an estimate based on M imputations to one based on $M = \infty$ number of imputations is approximately $1 + \gamma/M$ to 1, where γ is the rate of missing information. Assuming a fraction of 50% missing information an estimate based on $M = 3$ imputations has a standard error that is about 8% higher than one based on $M = \infty$, because $\sqrt{1 + 0.5/3} = 1.0801$. Schafer (1998) states that unless the fraction of missing information is higher than 50% there is little benefit in using more than 5 to 10 imputations.

See Kim et al. (2006) for more details on bias of the MI.

2.4.1 Multiple Imputation and the single imputation procedure: the ABB method

A basic issue of the MI is the single imputation method repeated M times. Theoretically, the method assumes the Y_{mis}' are M repetitions from the posterior predictive distribution of Y', each repetition being an independent drawing of the parameters and missing values under appropriate Bayesian models for the data and the posited response mechanism. In practice, three aspects of the imputation method have to be considered:

- if the underlying imputation model is explicit or implicit;
- if the underlying imputation model is ignorable or nonignorable;
- if the imputation methods is proper or not proper.

The first two concepts are typically dealt with in the single imputation approach as well. Commonly in the Official Statistic ignorable model are assumed, while we focus on the random hot deck method that falls in the method using implicit imputation modeling.

Here the concept of proper/not proper method is introduced. Imputation procedures that incorporate appropriate variability among the repetitions within a model (explicit or implicit, ignorable or nonignorable) are called proper (Rubin 1987). The reason for using proper imputation methods is that they properly reflect sampling variability when creating repeated imputations under a model, and as a result lead to valid inferences. For example, assume ignorable nonresponse so that respondents and nonrespondents with a common value auxiliary variable X (i.e. imputation cell) have Y values only randomly different from each other. Randomly drawing imputations for nonrespondents from matching respondents' Y values ignores some sampling variability. This variability arises from the fact that the sampled respondents' Y values at X randomly differ from the population of Y values at X. Properly reflecting this variability leads to repeated imputation inferences that are valid under the posited response mechanism. In particular Rubin and Schenker (1986) examined the hot deck procedure with MI. The imputation method assumes within the hot deck cells responses are missing randomly and the Y's are independent random variables with common mean and variance. For each unit having a missing value M values are imputed. The authors shown the standard hot deck procedure is not proper and variance with MI performs a variance underestimate. They proposed the Approximate Bayesian Bootstrap (ABB) for simple random sampling with hot deck imputation and MI method. Such technique can be viewed as a hot deck imputation method in the MI context (Kim and Fuller 2004; Kim et al. 2004).

Let us consider a collection of n units in the specific hot deck cell where there are n_r respondents and $n_{nr} = n - n_r$ nonrespondents. The ABB creates M ignorable repeated imputations as follows. For $i = 1, \ldots, M$ create n possible values of Y by first drawing n values at random with replacement from the n_r observed values of Y, and second drawing the n_{nr} missing values of Y at random with replacement from those n values. The drawing of n_{nr} missing values from a possible sample of n values rather than the observed sample of n_r values generates appropriate between imputation variability, at least assuming large simple random samples at X showing that is a proper method.

The ABB approximates the Bayesian Bootstrap by using a scaled multinomial distribution to approximate a Dirichlet distribution.

When the imputation cells cut across the sampling strata, unequal inclusion probabilities should be involved in the procedure. Nevertheless, no literature discusses the applications of ABB in this context. Some authors suggest (Brick et al. 2005) to disregard the unequal inclusion probabilities in the ABB. In the simulation below the Y values are drawn at random with equal inclusion probabilities.

3. Specialized results

A Monte Carlo simulation has been carried out for comparing the modified EDAGJK with bootstrap and MI. In the simulation a standard sampling strategy for the business survey has been implemented. The analysis of the results focuses on the statistical properties both with applicability of the methods in case of a complex survey sampling typically conducted by a National Statistical Institute.

3.1 The population and the sampling strategy

The simulation is based on the real data of the 2008 Italian enterprises belonging to the economic activity 162 according to the Statistical Classification of Economic Activities in the European Community NACE Rev.2 3-digit (number of units $N = 21,231$). This sub-population is surveyed in the Small and Medium Enterprises (SME) survey. The SME is a yearly survey investigating the profit-and-loss account of enterprises with less than 100 employed persons, as requested by SBS EU Council Regulation n. 58/97 (Eurostat, 2003) and n. 295/2008. The Italian target population of the SME survey is about 4.5 millions active enterprises.

The following target variables have been considered: Changes of inventory of finished and semifinished products (CIF); Purchase of commodities (PUC); Operating expenses for administration (OEX); Labour cost (LCO). The values of the four variables have been taken from the balance sheets (administrative data) for the whole population.

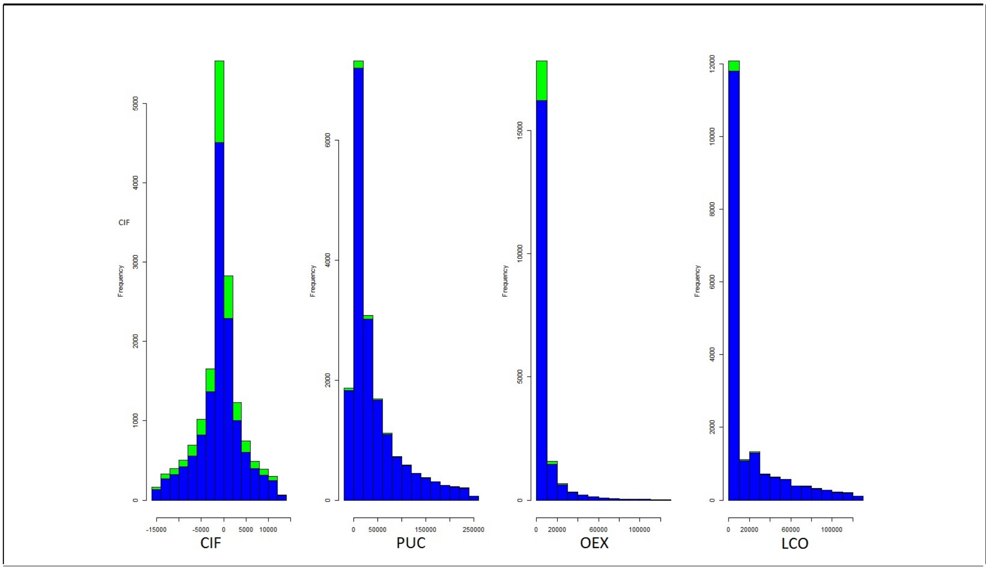
Table 1 gives some summary statistics.

Table 1 - Summary of the target variables

Variables	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
CIF	-869300	-3750	-130	-2422	1490	322700
PUC	-6631	5514	19710	41880	56230	247900
OEX	0	441	1601	6201	5294	126200
LCO	0	0	0	16700	24270	126500

Figure 1 shows that CIF variable has a symmetric distribution while especially OEX and LCO have highly skewed distributions.

Figure 1 - Distribution of the variables of interest and frequncy of non responses (light colour)



The simulation takes into account of a simplifyed version of the current sampling strategy used in the SME survey. A stratified simple random sampling design and a calibration estimator have been considered.

Strata are obtained by crossing the size classes and the regions according to the Nomenclature of territorial units for statistics NUTS 1 defined by EU. Hence, 20 strata are obtained as aggregation of the original strata of SME survey. The sample allocation in each stratum is taken from the allocation of 2008 SME survey. Table 2 shows the population and sample distribution in each stratum. The overall sample size is $n = 908$ enterprises.

The estimator calibrates the sampling weights to the number of enterprises and the number of employed persons at NACE Rev.2 4-digit, size class and NUTS 1 region.

The linear distance function (generalized regression estimator) is considered. Actually, the logit distance function is used in the SME survey, because it produces nonnegative calibrated weights. Nevertheless the logit distance has two drawbacks: the convergence is not guaranteed; it requires a time spending iterative procedure to obtain the calibrated weights. For these reasons linear distance function has been preferred in the simulation. We point out that the calibration estimators with linear and logit distance function converge asymptotically and the simulation results with the linear distance will be coherent with the ones obtained with the logit distance function.

The main task of the simulation study is to compare different methods of variance estimation for estimators of totals in a complex context, usual in the business survey, such as: stratified simple random sampling, imputation for item nonresponse and calibration estimator. The item nonresponses are imputed by means of random hot-deck procedure.

Besides this sampling strategy the simulation regards the variance of the Horvitz-Thompson (HT) estimator.

Table 2 - Population, sample distribution and missing rates (%) in the design strata

Strata (NUTS 1 region by size class)	Number of enterprises in the population	Number of enterprises allocated in the sample	Sampling rate	Missing rate by variable CIF	PUC	OEX	LCO
NORTH-WEST:(size class<8)	5241	107	0.02	16.75	1.28	7.52	1.97
NORTH-WEST:(9<size class<18)	425	39	0.09	14.59	1.41	3.29	0.00
NORTH-WEST:(19<size class<28)	83	17	0.20	28.92	2.41	8.43	0.00
NORTH-WEST:(size class>29)	38	15	0.39	13.16	5.26	5.26	0.00
NORTH-EAST:(size class<8)	5087	89	0.02	18.28	2.54	9.44	3.07
NORTH-EAST:(9<size class<18)	550	48	0.09	14.18	3.45	6.00	0.18
NORTH-EAST:(19<size class<28)	129	44	0.34	15.50	4.65	7.75	0.00
NORTH-EAST:(size class>29)	55	20	0.36	19.35	3.23	3.23	0.00
CENTER:(size class<8)	3699	145	0.04	19.60	1.14	9.22	1.57
CENTER:(9<size class<18)	297	34	0.11	19.53	2.69	5.05	0.00
CENTER:(19<size class<28)	63	35	0.56	19.35	3.23	3.23	0.00
CENTER:(size class>29)	31	17	0.55	20.79	2.60	10.43	2.78
SOUTH:(size class<8)	3386	128	0.04	17.61	1.14	5.11	0.00
SOUTH:(9<size class<18)	176	45	0.26	13.51	0.00	0.00	0.00
SOUTH:(19<size class<28)	37	23	0.62	26.67	0.00	0.00	0.00
SOUTH:(size class>29)	15	8	0.53	18.64	1.16	10.21	2.83
ISLANDS:(size class<8)	1803	50	0.03	13.48	0.00	3.37	1.12
ISLANDS:(9<size class<18)	89	26	0.29	4.76	0.00	0.00	0.00
ISLANDS:(19<size class<28)	21	16	0.76	4.76	0.00	0.00	0.00
ISLANDS:(size class>29)	6	2	0.33	16.67	16.67	16.67	16.67
		Missing rate	Average	18.36	1.88	8.73	2.19

3.2 Item nonresponse model

The item nonresponses have been generated seeking to reproduce the item nonresponse pattern of the business surveys.

The SME survey suffers from item nonresponses. Actually the survey has not flags for item nonresponses and the item nonresponses are denoted by zero values.

To find out when a zero value means a real zero or a missing value, the 2008 SME data have been linked with the 2008 administrative data; the zero SME values corresponding to a non zero value in the administrative data have been identified as missing values.

A regression tree model (rpart R package) has been applied for estimating the relationship between response propensity and outcome-related auxiliary variables known for the whole population. To create missing values, response indicators were assigned to the units within nonresponse cells defined by the regression tree. Within a given response cell, units were assigned at random to be missing or nonmissing at a specified rate. The mechanism generating missingness assumes that there is a uniform response probability within each cell. This is an usual assumption for the nonresponse model even though generally a more complex unknown item nonresponse model holds. When the real nonresponse model disagrees with the working model used for the imputation, the estimates are biased. However, the simulation is focused on the variance estimate and then we define an experimental context in which the point estimates are unbiased, for not creating confounding evidences.

Table 2 shows the strata missing rates (the last four columns) for the variables of interest. Then, three type rates of item nonresponse appear: high for the CIF variable with average equal to 18.36%, medium for the OEX variable with 8.73% and low for the PUC and LCO variables with about 2% of nonresponse rate. Figure 1 underlines that for CIF variable the missing rate increases when the frequency enlarges. For the other three variables the missing rate is concentrated on the smaller values of the variable.

We checked the unbiasedness of the estimator after the hot-deck imputation computing the empirical relative bias

$$RB(\hat{\vartheta}) = \frac{1}{C}\sum_{c=1}^{C}\frac{(\hat{\vartheta}_c - \theta)}{\theta} \tag{18}$$

being $\hat{\vartheta}_c$ the estimate from the sample c drawn according to the sampling design of section 3.1 and C the number of drawn samples. To obtain a nearly zero relative bias $C = 10,000$ samples have been selected.

Table 3 shows negligible bias for all estimates: e. g. the $RB(\hat{\vartheta})\%$ are lower than 1% except for the variable CIF when calibration with imputed data is considered (1.4%).

Table 3 - Relative bias($RB(\hat{\vartheta})\%$) of the estimators

Estimators	CIF	PUC	OEX	LCO
HT with imputation	0.72	0.49	0.94	-0.09
CALIBRATION with imputation	1.40	0.51	0.97	-0.10

3.3 Results of the Monte Carlo simulation

Several methods are compared in the simulation. Furthermore, the following reference variances

$$V(\hat{\vartheta}) = \sum_{c=1}^{10,000}\frac{(\hat{\vartheta}_c - \theta)^2}{10,000}, \tag{19}$$

hereinafter denoted as empirical or Monte Carlo variances, are computed.

For the HT estimator are considered the:

- unbiased variance estimator (Wolter 2007) denoted as STANDARD method;
- EDAGJK according to Kott (Kott 2001);
- EDAGJK.I: the modified EDAGJK (section 2.2 using the replicate weights given in the (3);
- BOOTSTRAP.I: bootstrap variance methods under imputation (Shao and Sitter 1996);
- MI: using the Approximate Bayesian Bootstrap (ABB) (Kim et al. 2004, Brick et al. 2005).

For the calibration estimator have been compared the:

- TAYLOR variance estimator;
- EDAGJK.HT computed according to (4);
- EDAGJK.CAL computed according to (5);
- EDAGJK.HT.I: the modified EDAGJK (section 2.2) based on the correction factor (4);
- EDAGJK.CAL.I: the modified EDAGJK (section 2.2) based on the correction factor (5);
- BOOTSTRAP.I: bootstrap variance methods under imputation (Shao and Sitter 1996);
- MI: using the ABB.

Note that the STANDARD, EDAGJK, TAYLOR, EDAGJK.HT and EDAGJK.CAL, do not properly take into account the imputation correction.

The complete simulation has implemented twelve variance methods by four variables. The imputation procedure is performed and for seven variance estimators imputation adjustement is carried out as well. Then, computational issues led us to choose 1,000 replications in performing the variance estimates for each method.

The accuracy of the variance estimates is measured with following summary statistics:

- The Relative (percentage) Bias of Variance estimation

$$RB[v(\hat{\vartheta})]\% = 100 \times \frac{\bar{v}(\hat{\vartheta}) - V(\hat{\vartheta})}{V(\hat{\vartheta})}. \tag{20}$$

- The Relative (percentage) Root Mean Square Error of Variance estimation

$$RRMSE[v(\hat{\vartheta})]\% = 100 \times \sqrt{\frac{\frac{1}{1{,}000}\sum_{c=1}^{1{,}000}\left[v(\hat{\vartheta}_c) - V(\hat{\vartheta})\right]^2}{V(\hat{\vartheta})^2}}. \tag{21}$$

- The Coverage of the Confidence Interval (percentage), that is the percentage of intervals including θ, based on the nominal 95 % confidence intervals computed for each of 1,000 simulations. We used the normal distribution as approximation of the t distribution
  $CCI[v(\hat{\vartheta})]\% = \frac{100}{1{,}000}\sum_{c=1}^{1{,}000}\delta_c$ where $\delta_c = 1$ if $\theta \in \left(\hat{\vartheta}_c \pm 1.96\sqrt{v(\hat{\vartheta}_c)}\right)$ and $\delta_c = 0$ otherwise.
- The Lower Error Rate and Upper Error Rate
  $LER[v(\hat{\vartheta})]\% = 100 \times \frac{1}{1{,}000}$(number of samples with $\theta < -1.96\sqrt{v(\hat{\vartheta}_c)}$),
  $UER[v(\hat{\vartheta})]\% = 100 \times \frac{1}{1{,}000}$(number of samples with $\theta > +1.96\sqrt{v(\hat{\vartheta}_c)}$).

Table 4 shows that for the variables with large nonresponse rate (CIF and OEX), the methods that do not take properly into account the imputation process such as STANDARD, EDAGJK, TAYLOR, EDAGJK.HT and EDAGJK.CAL produce large downward biased variance estimates. The result was definitely expected.

Furthermore for the OEX variable we observe a very large variability with the $RRMSE[v(\hat{\vartheta})]\%$ over than 177% for all the methods. The evidence is explained by the positive skew distribution of this variable (figure 2).

The scatterplot of the 10,000 variance estimates versus the corresponding HT estimates (figure 3) shows for the OEX variable two separate clouds, being the highest one around of size 200. This is due to one extreme value within the stratum NORTH-WEST:(size class<8) with the 0.02 sampling rate. For this stratum the expected percentage contribution to the overall variance is around 85%. Furthermore, the stratum variance when the extreme value is included is about 5 times the stratum variance when the extreme value is not included in the sample.

The presence of rare extreme values is typical in the business surveys and then it is interesting to study the behaviour of the estimators in this critical context.

In the following the main comments are focused on CIF variable, because it has the highest missing rate. As concerns the HT estimator EDAGJK.I and bootstrap methods, they produce $RB[v(\hat{\vartheta})]\%$ around 7% but bootstrap has a smaller $RRMSE[v(\hat{\vartheta})]\%$ than EDAGJK.I. MI has the smallest $RB[v(\hat{\vartheta})]\%$ with the drawback to be negative.

Table 4 - Relative Bias ($RB[v(\hat{\vartheta})]$) and Relative Root Means Square Error ($RRMSE[v(\hat{\vartheta})]$) of the variance estimators with imputed data

	H-T estimator							
	$RB[v(\hat{\vartheta})]\%$				$RRMSE[v(\hat{\vartheta})]\%$			
Variance Estimator	CIF	PUC	OEX	LCO	CIF	PUC	OEX	LCO
STANDARD	-27.15	-3.37	-8.96	-6.40	53.43	38.54	204.64	15.41
EDAGJK	-21.26	3.11	-8.11	-0.30	57.94	45.90	178.76	31.26
EDAGJK.I	7.43	7.44	6.77	3.99	71.35	47.65	209.31	32.61
BOOTSTRAP.I	7.95	5.57	21.40	2.81	64.53	42.99	261.52	21.90
MI	-2.83	2.36	-1.82	-3.08	66.74	40.96	205.00	15.58
	Calibration estimator							
	$RB[v(\hat{\vartheta})]\%$				$RRMSE[v(\hat{\vartheta})]\%$			
TAYLOR	-28.48	-6.21	-9.82	-7.55	53.90	43.93	227.47	23.82
EDAGJK.HT	-19.76	3.60	-5.73	2.95	59.27	51.12	195.25	36.73
EDAGJK.CAL	-18.24	6.79	-6.56	4.98	62.15	55.04	186.05	40.89
EDAGJK.HT.I	9.48	8.01	9.15	6.61	73.30	53.02	225.78	38.60
EDAGJK.CAL.I	11.51	11.31	8.23	8.75	77.73	56.85	215.29	42.76
BOOTSTRAP.I	5.71	6.57	14.43	4.79	64.38	48.84	244.40	27.39
MI	-9.05	-4.46	-15.34	0.17	57.24	42.46	177.88	22.42

In case of calibration estimator, bootstrap outperforms the methods that consider specifically the imputation in terms of $RB[v(\hat{\vartheta})]\%$. Nevertheless the modified EDAGJK methods produce positive and not so large $RB[v(\hat{\vartheta})]\%$. MI has negative bias, but outperfoms the other estimators in terms of $RRMSE[v(\hat{\vartheta})]\%$. The table 4 shows that EDAGJK.HT is slightly better than EDAGJK.CAL at least for the CIF variable.

Table 5 shows the coverage of the confidance interval. The methods ignoring that many values are imputed have a strong reduction of the coverage rates. MI does not show good performaces, at least for the CIF variable, while a small decreasing of coverage is observed for the rest of the resampling methods. The modified EDAGJK techniques and bootstrap are essentially equivalent. For the calibration estimator, EDAGJK.CAL.I seems slightly better than EDAGJK.HT.I and bootstrap. That occurs because of a larger $RB[v(\hat{\vartheta})]\%$.

Table 5 - Coverage of the Confidence Interval($CCI[v(\hat{\vartheta})]$), the Lower and Upper Error Rate ($LER[v(\hat{\vartheta})]$, $UER[v(\hat{\vartheta})]$) with imputed data

	H-T estimator											
	$CCI[v(\hat{\vartheta})]\%$				$LER[v(\hat{\vartheta})]\%$				$UER[v(\hat{\vartheta})]\%$			
Variance Estimator	CIF	PUC	OEX	LCO	CIF	PUC	OEX	LCO	CIF	PUC	OEX	LCO
STANDARD	88.40	94.50	90.40	94.40	7.50	1.20	1.00	1.40	4.10	4.30	8.60	4.20
EDAGJK	86.20	93.00	88.20	94.00	6.80	4.70	7.70	3.40	7.00	2.30	4.10	2.60
EDAGJK.I	91.10	93.70	89.90	94.40	4.20	4.20	6.90	3.10	4.70	2.10	3.20	2.50
BOOTSTRAP.I	92.30	93.80	90.30	95.90	2.90	4.20	7.10	2.30	4.80	2.00	2.60	1.80
MI	89.90	94.40	90.60	95.50	4.50	3.90	7.20	2.50	5.60	1.70	2.20	2.00
	Calibration estimator											
	$CCI[v(\hat{\vartheta})]\%$				$LER[v(\hat{\vartheta})]\%$				$UER[v(\hat{\vartheta})]\%$			
TAYLOR	88.20	94.20	90.20	93.40	7.70	1.80	0.90	2.00	4.10	4.00	8.90	4.60
EDAGJK.HT	86.20	92.70	89.40	93.60	6.50	5.30	7.00	3.00	7.30	2.00	3.60	3.40
EDAGJK.CAL	86.76	93.88	89.87	92.98	6.32	4.61	7.12	3.51	6.92	1.50	3.01	3.51
EDAGJK.HT.I	91.80	92.90	91.20	93.70	3.50	5.20	6.30	3.00	4.70	1.90	2.50	3.30
EDAGJK.CAL.I	92.38	94.18	91.57	93.68	3.71	4.41	5.82	3.21	3.91	1.40	2.61	3.11
BOOTSTRAP.I	92.00	93.72	90.08	95.65	3.24	4.66	6.88	1.92	4.76	1.62	3.04	2.43
MI	90.50	92.20	88.20	93.50	4.00	6.00	8.40	3.30	5.50	1.80	3.40	3.20

Figure 2 - Distribution of the 10,000 HT estimates

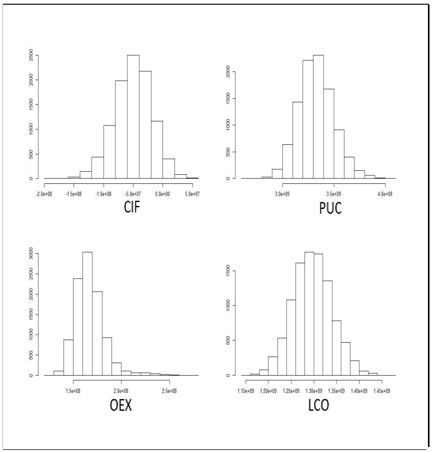
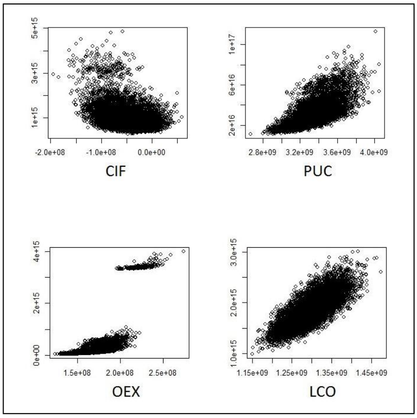


Figure 3 - Scatterplot of the HT estimates versus standard variance estimates



4. Conclusions

Many statistical surveys carried out by National Statistical Institutes are, generally, defined by a large sample drawn according to a complex sampling design. Unit and item non responses increase the trouble to make inference.

The paper investigates three variance estimators taking into account the item non responses when random hot deck imputation has been performed. Two of the three methods are the standard bootstrap and the MI, while the third one is a new variance estimator. The proposed method combines the EDAGJK technique proposed by Kott with the adjusted jackknife proposed by Rao and Shao. The reasons leading to new estimator is the good compromise among theoretical properties and practical aspects. In particular EDAGJK produces an unbiased estimator (for complete data set), it is easy to implement and not computer intensive. Furthermore, the adjusted jackknife does not require replications of the imputation procedure.

These features are quite appealing especially in a National Statistical Institute (NSI), where data production based on large data sets must be automatized as much as possible.

The three methods have been compared by means of a Monte Carlo simulation based on real business data and a sampling strategy resembling to the Small and Medium Entrerpise survey conducted by the Italian Statistical Institute. The simulation results show that the modified EDAGJK[4] with Rao and Shao adjustment produces nearly unbiased variance estimates and it works well with respect to the two benchmarking methods in terms of accuracy and coverage of coinfidance interval. Nevertheless the method is less computational demanding than bootstrap and it does not require an increasing of complexity of the data production process as for MI.

The paper show that for variable with an hight level of imputation the standard methods of variance estimation deeply under-estimates the true variance, a best practice for a NSI should be to consider the level of item non response for each variables and to performance

The empirical results shows that for variables with an high level of imputation rate, the standard methods of variance estimation deeply under-estimate the true variance. Then a

4 An R function implementing the modified EDAGJK is available in the Deliverable 6.1 of the BLUE-Enterprise and Trade Statistics project (Blue-ETS 2013).

best practice for NSIs should be to make a screening, for the main variables of interest of each business survey, of the item non response rates and to adopt valid variance estimation methods for the variables affected by the highest (eg. >10%) item non response rates.

The information about the item non-response rates shuld be also disseminated to the external users. In fact, if the research institute releases a standard file with imputation flag variable and the replicate weights, every users can compute the variance estimates for every kind of unplanned domains of interest by a simple formula.

References

Blue-ETS (2013): Deliverable 6.1. Best practice recommendations on variance estimation and small area estimation in business surveys. Computer Codes. URL http://www.blue-ets.istat.it/fileadmin/deliverables/Deliverable6.1.pdf

Brick, J. M., Jones, M. E., Kalton, G. and Valliant, R. (2005): Variance Estimation with Hot Deck Imputation: A simulation Study of Three Methods. *Survey Methodology*, 31, pp. 151-159.

Burns, R. M. (1990): Multiple and Replicate Item Imputation in a Complex Sample Survey. *Proceedings Sixth Annual Res. Conf.*, pp. 655-665, Washington, DC: U.S. Bureau of the Census.

Chen, J. and Shao, J. (2001): Jackknife Variance Estimation for Nearest-Neighbour Imputation. *Journal of the American Statistical Association*, 95, pp. 260-269.

Di Zio, M., Falorsi, S., Guarnera, U., Luzi, O. and Righi, P. (2008): Variance Estimation in Presence of Imputation: an Application to an Istat Survey Data. *Proceedings of the Sec. on Survey Res. Meth. European Conference on Quality in Official Statistics*. URL http://q2008.istat.it/sessions/paper/17DiZio.pdf

Efron, B. (1994): Missing Data, Imputation and the Bootstrap. *Journal of the American Statistical Association*, 89, pp. 463 - 479.

Kalton, G. and Kasprzyk, D. (1986): The Treatment of Missing Survey Data. *Survey Methodology*, 12, pp. 1-16.

Kim, J. K., Brick, J. M., Fuller, W. A. and Kalton, G. (2006): On the Bias of the Multiple-Imputation Variance Estimator in Survey Sampling. *J. R. Statist. Soc. B*, 68, pp.509-521.

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

Kott, P. (1998): Using the Delete-a-Group Jackknife Variance Estimator in NASS Surveys. *Nass research report 98-01* (revised 2001), NASS. URL http://www.nass.usda.gov/research/reports/RRGJ7.pdf

Kott, P. (2001): Delete-a-Group Jackknife. *Journal of Official Statistics*, 17, pp. 521-526.

Kott, P. (2006): Delete-a-Group Variance Estimation for the General Regression Estimator Under Poisson Sampling. *Journal of Official Statistics*, 22, pp. 759-67.

Little, R. J. A. and Rubin, D. B. (2002): *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics, Wiley.

Miller, D. and Kott, P. (2011): Using the DAG Jackknife to Measure the Variance of an Estimator in the Presence of Item Nonresponse. *JSM Proceedings, Statistical Computing Section*. Alexandria, VA: Am. Statist. Assoc.

Rao, J. N. K. (1996): On Variance Estimation with Imputed Survey Data. *Journal of the American Statistical Association*, 91, pp. 499-506.

Rao, J. N. K. and Shao, S. (1992): Jackknife Variance Estimation with Survey Data Under Hot Deck Imputation. *Biometrika*, 79, pp. 811-822.

Rao, J. N. K. and Shao, S. (1999): Modified Balanced Repeated Replication for Complex Survey Data. *Biometrika*, 86, pp. 403-415.

Rao, J. N. K. and Sitter, R. R. (1995): Variance Estimation under Two-Phase Sampling with application to Imputation for Missing Data. *Biometrika*, 82, pp. 453-460.

Rao, J. N. K. and Wu, C. F. J. (1988): Resampling Inference with Complex Survey Data. *Journal of the American Statistical Association*, 83, pp. 231-241.

Raghunathan, T. E., Lepkowski, J.M., Van Hoewyk, J. and Solenberger, P.: A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. *Survey Methodology*, June 2001 27, pp. 8595

Rubin, D. B. (1978): Multiple Imputation in Sample Surveys. *Proceedings Survey Res. Meth. Sec.*, Am. Statist. Assoc., pp. 20-34.

Rubin, D. B. (1987): *Multiple Imputation for Nonresponse in Surveys*. Wiley Series in Probability and Statistics, Wiley.

Rubin, D. B. and Schenker, N. (1986): Multiple Imputation for Interval Estimation for Simple Random Samples with Ignorable Nonresponse. *Journal of the American Statistical Association*, 81, pp. 366-374.

Rust, K. (1985): Variance Estimation for Complex Estimators in Sample Surveys. *Journal of Official Statistics*, 1, pp. 381-397.

Rust, K. (1986): Efficient Formation of Replicates for Replicated Variance Estimation. *Proceedings Survey Res. Meth. Sec.*, Am. Statist. Assoc., pp. 81-87.

Rust, K. and Rao, J. N. K. (1996): Variance Estimation for Complex Surveys Using Replication Techniques. *Statistical Methods in Medical Research*, 5, pp. 283-310.

Saigo, H., Shao, J. and Sitter, R. R. (2001): A Repeated Half-Sample Bootstrap and Balanced Repeated Replications for Randomly Imputed Data. *Survey Methodology*, 27, pp. 189-196.

Saigo, H. and Sitter, R. R. (2005): Jackknife Variance Estimator with Reimputation for Randomly Imputed Survey Data. *Statistics and Probability Letters*, 73, pp. 321-331.

Schafer, J. L. (1998): Multiple Imputation: a Primer. *Statistical Methods in Medical Research*,8, pp. 3-15.

Shao, J. (2003): Impact of the Bootstrap on Sample Surveys. *Statistical Science*, 18, pp. 191-198.

Shao, J. and Sitter, R. R. (1996): Bootstrap for Imputed Survey Data. *Journal of the American Statistical Association*, 91, pp. 1278-1288.

Shao, J. and Tang, Q. (2011): Random Group Variance Estimators for Survey Data with Random Hot Deck Imputation. *Journal of Official Statistics*, 27, pp. 507-526.

Shao, J. and Tu, D. (1995): *The Jacknife and Bootstrap*. Springer-Verlag GmbH.

Skinner, C. J. and Rao, J. N. K. (2002): Jackknife Variance Estimation for Multivariate Statistics under Hot-Deck Imputation from Common Donor. *Journal of Statistical Planning and Inference*, 102, pp. 149-167.

Särndal, C. E. (1992): Methods for Estimating the precision of Survey Estimates when Imputation has been used. *Survey Methodology*, 18, pp. 241-252.

Valliant, R., Brick, M. J. and Dever, J. (2008): Weight Adjustments for the Grouped Jackknife Variance Estimator. *Journal of Official Statistics*, 24, pp. 469-488.

Wolter, K. (2007): *Introduction to Variance Estimation*. Springer London, Limited.

Yung, W. and Rao, J. N. K. (2000): Jackknife Variance Estimation Under Imputation for Estimators Using Poststratification Information. *Journal of the American Statistical Association*, 95, pp. 903-915.