WP. 36
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Ottawa, Canada, 16-18 May 2005)

Topic (iv): New and emerging methods, including automation through machine learning, imputation, evaluation of methods

INTEGRATED MODELLING APPROACH TO IMPUTATION AND DISCUSSION ON IMPUTATION VARIANCE

Supporting Paper

Submitted by Statistics Finland[1]

I. INTRODUCTION

During recent years the imputation strategy in Statistics Finland has been concretised so that we can speak about an integrated modelling approach (IMAI) to this kind of datacleaning (see Laaksonen 2000&2002a&2003, Chambers et al 2001, Piela and Laaksonen 2001). The strategy has been under consideration and consequently further developed in the two big European Union research projects, that is, in Euredit and Dacseis.

The Euredit project concentrated on developing and comparing different methods for editing and imputation. As far as imputation is concerned, this project paid most attention to the bias, using the special performance criteria developed during the project (Chambers 2001). These criteria were used when comparing the performance of a number of imputation methods that were divided to the two major groups: (i) standard methods and (ii) new methods (that means neural nets methods). In fact, some standard methods also include many new elements, and respectively some so-called new methods had been used before the Euredit project. However, most neural nets methods, especially self-organised maps (SOM), correlation matrix memory (CMM) and support vector machine (SVM) were implemented the first time for imputation purposes during the project. The same was concerned some editing techniques too, but this paper only rarely discusses editing questions.

Euredit thus tried to make imputations so that those preservation measures (concerning preservation of 'true' values) performed as well as possible. Since we afterwards thus knew these 'true' values, the exact performance checks were possible to do. Awkwardly, the different performance criteria gave often different conclusions when comparing a number of methods and different estimates, no unambiguous conclusion could be drawn, although some methods were superior to some others with most criteria. On the other hand, the project did not try to estimate in advance any precision measure for estimates when some data were imputed, that is, the Euredit project did not paid attention to imputation variance that would be a good instrument for a user when handling partially imputed data in real life when the real values cannot be known.

[1] Prepared by Seppo Laaksonen, email: Seppo.Laaksonen@Stat.Fi or Seppo.Laaksonen@Helsinki.Fi

The Dacseis project, at contrast, just concentrated on variance estimation including some experiments with imputed data too. The imputation methods as well as the missingness patterns were however simpler in Dacseis than in Euredit, and hence we have no strong answers to many big questions such as:
- In which case we should use imputation?
- Which imputation method is best if the different alternatives give contradictory results?
- How to obtain (estimate) a solid measure for the uncertainty of the estimate?
- What to do if the value of an uncertainty indicator is 'too big'?

II. INTEGRATED MODELLING APPROACH TO IMPUTATION (IMAI)

In this paper, we cannot give the comprehensive answers to the above topical questions either but we try to explain the principles of the IMAI approach. This approach is rather general and can be used both in the context of 'standard methods' and of new methods. The approach also works with multiple imputation (MI), although some IMAI specifications are only available for single imputation (SI). I here want to point out, that this IMAI approach does not see MI as any special imputation method (e.g. Solas has a different approach), but preferably as a potential tool for estimating the uncertainty of imputations (imputation variance). Thus we have first to concentrate on finding the best possible (single) imputation technique, as done in Euredit, and when we are satisfied with this (or several competitive techniques), we naturally have to try to provide a reliable uncertainty measure. For this step, MI can be a rational solution if we can find a good stochastic tool for this duplication (or make imputation properly as said by Rubin 1987). Thus I see that the preservation of real values (although it is of type 'black box') is most important in imputations.

The IMAI approach is based on the following five steps (the strategy is analogous in editing)[2]:

A. Selection of training data and auxiliary variables for step B
B. Construction of imputation model
C. Choice of criteria for imputation
D. Imputation task itself.
E. Calculation of the uncertainty for the estimates (incl. imputation variance)

Some principal notes to each of these steps:

A: There should be a maximal potentiality of auxiliary variables with non-missing values or such values which have been considered as non-missing (Laaksonen 2002b). The latter point can mean for example, that some values of the auxiliary variable may be imputed earlier (sequential imputation) or that there can be a special category for missingness in the case of a categorical auxiliary variable. In this step, an 'imputeur' has also to decide whether to divide that the full data into several sub-data within those to apply the next steps or alternatively work with one complete data set. This decision may be made also using automatic tools such as are available in tree-based methods (a useful tool is software WAID, see Chambers et al 2001) or in SOM, for example.

B: The two alternative target variables may be used: (i) the target variable with missing values or (ii) a missingness indicator of the target variable. Explanatory variables without missing values only are accepted. A model for each particular case may be of a whatever type, thus parametric or non-parametric, the model may be estimated from data or 'logically deducted,' that is, the model is constructed based on prior or plausible information. The key target in modelling is high predictability. In case of approach (ii) the most common model is logistic regression that exploits *logit* link, but naturally the other links (*probit, log-log, complementary log-log*) may be used although the results obviously will not differ very much with each other. At contrast, approach (i) may cover a lot of alternative models and their specifications (incl. different distribution assumptions and link functions in parametric models, and non-parametric approaches using for example tree-based methods and SOM methods).

[2] Naturally, before that the missingness pattern of the data must have been done and analyzed, and then decided which values will be imputed.

C: The criteria for imputation are of two types: (i) assumptions for direct predictability or (ii) metrics for nearness. Typically, such a metrics is based on a Euclidean distance measure or other model-external solutions, often using such auxiliary variables which are not used in a model. Alternatively, the metrics can be taken from model results (with or without noise term) as in the case of the 'regression based nearest neighbour' (RBNN) technique (Laaksonen 2000&2002a&2003).

D: If the modelled values (predicted with or without noise term) are used as the imputed values, we speak about '*model-donor*' methods, whereas if a model and a metrics have been used to find a good donor from whom an imputed value has been borrowed, we speak about '*real-donor*' methods. Note that this technique may be used for finding a good observed residual (noise term) too (see Laaksonen 2002a). Note thus also that a noise term may be determined in different manners such as assuming that it follows a certain theoretical distribution (normal, log-normal, both often useful to truncate) and it may be drawn from the set of observed residuals.

For example, in the case of linear regression model we thus have estimated the parameters of this model. The model may be specified very simply, including only the intercept term. If we use this intercept as each imputed value, the method corresponds to 'mean imputation.' If the only explanatory variable in the model is a categorical variable, and we forget the noise term, the method corresponds to 'cell-mean imputation.' If we add the noise term, and apply again a model-donor approach, the imputed values are category-intercepts plus random noises. Here the noises may be from our anticipated distribution, typically from a normal distribution with zero mean and with the MSE variance (MSE = mean square error of the estimated model). Usually, the model is more complex but the technique is analogous to model-donor methods. On the other hand, the estimated values (with or without noise term) of the same model may be used when searching for nearest neighbours for missing values (RBNN), that leads to real-donor techniques. If using a theoretical continuous distribution (such as normal distribution) behind the noise terms, there can be problems with such outliers that are not plausible. Hence it is often good to truncate this distribution, or to use observed residuals.

E. If a single imputation strategy includes a stochastic element for example via random noise term or if probabilities have been used in drawing a real donor (and this as such or together with additional options may be considered to provide proper imputation in the sense of Rubin 1987), we may quite straightforwardly estimate the MI-based variance for each estimate. Thus we duplicate the same procedure several times (often much more than 5-7 imputed values and completed data sets respectively are needed as also the Dacseis simulation experiments show). If the imputation strategy is not based on stochastic elements, we may try to calculate imputation variance using SI techniques (this is much discussed in the Imputation Bulletins of Statistics Canada for example). This work may be difficult in skewed distributions that is often the case in business surveys. Hence it is best to make all efforts for inquiring an as correct as possible value from a large business in particular than to impute such values badly.

There are several specifications for IMAI methods in Statistics Finland, most of them have been done using SAS. However, we have not yet provided any software. This could be possible in the future, rather so that some collaboration between other institutions would have been exploited. For some applications IMAI may be very easily applied, especially if no noise term has been used in a model. The main problem for a user is to decide the optimal options for each particular step of the IMAI strategy. It is very advantageous to look at data, and to test several model specifications. Finally, the options for step D depend much on the success in modelling, and on the users' requirements for the imputed data.

The Euredit experiments give some understanding both on good and not-so-good options. We had in some sense a competition between different methods and their specifications in Euredit. The six different data sets were used in these tests for imputation. Each tester worked first with the so-called *development data* in which case the true values were available. When a reasonable application was done, he/she removed to work with another data set that was similar to some extent but often not very much but the

structure in both data sets was the same. In this case, anyone did not know true values but the co-ordinator evaluated the imputation results using about 30 different criteria (Chambers 2001).

Next, the summary analysis of these different criteria were done and in some sense the methods and their specifications were ranked. IMAI participated in this competition for the four data sets, that is, UK Census, Danish income study, UK annual business inquiry and German socio-economic household panel (GSOEP). Its success was very good with each data, always IMAI was well ranked. Note that in most cases we could not say absolutely which method was a winner, because the results were partially contradictory (e.g. if a user would focus on aggregate results, not distributional aspects, a winner often changed). However, IMAI was in all criteria best with the GSOEP. The experiment results are available on the web, and so you can do your own interpretation on the results if you do not believe me. It is fair to add that since IMAI and some other methods as well are so much dependent on the success in imputation model, and since this may require much time or at least the reasonable understanding of the data and the phenomena behind the data, this technique is not always very productive and may be risky if the model builder is not competent. But can we make imputations to a new data set ever automatically? I do not believe it.

III. IMPUTATION VARIANCE

Imputations have been used in some way always when editing statistical data although this term has not been used in all such contexts. Some methods have not been either explicitly documented, they are subjective, or based on good guesses of a data editor. Naturally such simple methods as *last value carried forward* and *cell-based mean/median/mode imputation* have been used objectively as well, but the decision to use this method is subjective.

The research of imputation methods flourished from the late 1980s onwards, I suppose, and this trend is continuing. Most attention in developing this area has been paid to reduce for the bias due to missingness that is worsening in survey data. The advantage of imputations has been met especially in the case of metric variables of business and other economic surveys. Also the technique has been used in censuses concerning both metric and categorical variables as we made in Euredit exercises. Much less imputations have been used for attitudes or analogous subjective variables. The reason not to use imputation lies in the three main factors in my opinion: (i) the staff is not very competent to perform these operations, (ii) the missingness is not any dramatic problem, (iii) it often very difficult to construct a reasonably well fitting imputation model in some cases and hence provide reliable imputations.

While the use of imputations has been increasing in NSI's and other institutions, it is not clear for me how the data (and statistics) with imputed values have been assessed from the quality point of view. From the recent literature we find a lot of papers which give theoretical and also partially practical advice to this basic question. The term *'imputation variance'* has been introduced into use maybe 20 years ago. Consequently we can find a number of papers which examine this area. A good (theoretical) conclusion is available from the Dacseis project (Berger et al 2004). I have collected many other papers in the attached bibliography[3]. Multiple imputation approaches are not well covered in my list, however. A reason for this is that I do not know any practical exercise from European NSI's which has based on this approach (if I am wrong, please inform me). Naturally, a lot of tests have been done using NSI data, also the Dacseis in its simulations.

[3] When looking even at the titles of the papers, you can find which types of methods have been used for estimating imputation variance. Although there are several methods, the area has not been examined yet well, since in most exercises the imputation methods behind variance estimation are rather simple. For example, more innovations and implementations are needed in the case of such methods as RBNN, neural nets (SOM, CMM, SVM) and AID. I also want to notice that the different imputation variance methods may give quite different values. To which of these should a user believe?

An interesting question in this meeting would be to compare the experience of NSI's in the use of imputations. What is the current strategy? And especially, how much imputation variances and other quality measures are exploited in NSI's. And should we go forward to MI following Rubin's approach?

BIBLIOGRAPHY

Berger, Y. , Bjornstad, J., Zhang, L-C., Skinner, C. (2004). Imputation and Non-response. Dacseis Deliverable 11.1. 59 pages. www.dacseis.de/delieverables.

Chambers, R. (2001). *Euredit project. WP3: Interim report on evaluation criteria for statistical editing and imputation*. Euredit Report No. T001.04. University of Southampton. Available on the Euredit website.

Chambers, R.L., Hoogland, J., Laaksonen, S., Mesa, D.M., Pannekoek, J., Piela, P., Tsai, P. and de Waal, T. (2001). *The AUTIMP-project: Evaluation of Imputation Software*. Research Paper 0122. Statistics Netherlands.

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

Deville, J.C. and Särndal, C.E. (1994) Variance estimation for the regression imputed Horvitz-Thompson estimator. *Journal of Official Statistics***, 10**, 381-394**.**

Fay, R.E. (1996). Alternative paradigms for the analysis of imputed survey data. *Journal of the American Statistical Association*, **74**, 269-277.

Laaksonen, S. (2000). Regression-Based Nearest Neighbour Hot Decking. *Computational Statistics* 15, 1, 65-71.

Laaksonen, S. (2002a). Traditional and New Methods for Imputation. *Journal of Statistics in Transition*. December.

Laaksonen, S. (2002b). Need for High Level Auxiliary Data Service for Improving the Quality of Editing and Imputation. *Paper for the UNECE Work Session on Data Editing in Helsinki, 27-29 May*. Available, among others, on the UNECE website: www.unece.org.

Laaksonen, S. (2003). Alternative Imputation Techniques For Complex Metric Variables. *The Journal of Applied Statistics*, 1006-1021.

Lee, H., Rancourt, E. and Särndal, C.E. (2002). Variance estimation from survey data under single imputation. In R.M. Groves, P.A. Dillman, J.L. Eltinge and and R.J.A. Little eds. *Survey Nonresponse*, New York: Wiley.

Piela, P. and Laaksonen, S. (2001). Automatic Interaction Detection for Imputation – Tests with the WAID Software Package. *Conference Proceedings of the Federal Committee on Statistical Methodology*, Washington, November.

Rao, J.N.K. (1996). On variance estimation with imputed survey data. *Journal of the American Statistical Association*, **91**, 499-506.

Rao, J.N.K. and Shao, J. (1992). Jacknife variance estimation with survey data under hot deck imputation. *Biometrika*, **82**, 453-460.

Rao, J.N.K. and Shao, J. (1999) .Modified balanced repeated replication for complex survey data. *Biometrika*, **86**, 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**, 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**, 231-241.

Rubin D.B. (1987). *Multiple Imputation for Nonresponse in Surveys.* John Wiley & Sons Inc

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 Steel, P. (1999). Variance estimation with composite imputation and non-negligible sampling fractions. *Journal of the American Statistical Association*, **94**, 254-265.

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

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

Sitter, R. and Rao, J.N.K. (1997). Imputation for missing values and corresponding variance estimation. *Canadian Journal of Statistics*, **25**, 61-75.

Skinner, C.J. and Rao, J.N.K. (2002). Jacknife estimation for multivariate statistics under hot deck imputation from common donors. *Journal of Statistical Planning and Inference*,**102**,149-167.

SOLAS (2001). *Imputation User Reference Manual.* Statistical Solutions.

Yung, W. and Rao, J.N.K. (2000) Jacknife variance estimation under imputation for estimators using poststratification information. *Journal of the American Statistical Association*, 95, 903-915.

Euredit Project Website: http://www.cs.york.ac.uk/euredit/

Dacseis Project Website: http://www.dacseis.de/deliverables/

ANNEX

On Imputation Methods and Imputation Uncertainty in European National Statistical Institutes (NSI's) based on a small-scale survey in Spring 2005

When writing my supporting paper I observed that I very little know about practices in NSI's as far as the use of imputation methods is concerned and especially on how the imputed values have been taken into account in quality reports and in information for user's . Hence I performed a small-scale survey, and submitted it a number of NSI's in Europe. The response rate was not very high, less than 50%, but some light to my information lack these results will give. This is a short summary of these responses. I promised not to tell anything at a NSI level.

The five simple questions only were inquired. Next, the key points of each answer are given, and after that I present some summary points.

1. Mention the *three* important surveys or censuses in which you are using *imputation methods* (if possible choose such examples that exploit different methods):

(i) Business surveys/statistics:
- Short term business surveys,
- European Survey on ICT usage in households
- Material prices in construction
- Turnover index in construction
- Output prices in construction
- Foreign trade
- Panel surveys, e.g. price index and trade barometers
- Annual Industrial Survey with surveyed unit the enterprise belonging to manufacturing
- Survey on Employment and Wages
- Monthly survey on earnings (some earnings components)
- Monthly survey on turnover and new orders in industry
- Cost Structure Surveys
- Quarterly Survey on Services
- Structural business statistics (imputation of non-response of large enterprises)
- Structural business statistics (imputation of inconsistent variables in automated editing)
- Short term statistics in order to calculate and estimate values for selecting implausible records for interactive editing.

(ii) Social surveys/statistics:
- Labor Force Survey
- EU-SILC (European Survey for Income and Living conditions) (quantitative variables like income)
- Survey on Income and Living Conditions
- Education among immigrates
- Household-based time use survey (imputation for some household members with missing items)

(iii) Others:
- Population Census
- Demographic Statistics (based on register data)
- Occupation and socioeconomic status in register based census

2. What is the most challenging method in each example and for which type of variable(s) it is used?

-System DIA software for detection and automatic imputation (qualitative data)
-Specific program developed, in general, in SAS software, to impute quantitative data. Imputation methods: ratio imputation and hot-deck imputation are the most used
-Hot deck (problems of finding appropriate sorts)

-Regression models for income components
-Nearest neighbor based on distance functions
-The average price growth of enterprises providing the same material is implemented.
-The average sector growth (in the CC sense) is implemented.
- Models e.g. regression
- Last value carry forward
- The method of class mean imputation: the surveyed units are divided into classes according to values of the auxiliary variables being used for the imputation. Within each imputation class the respondent class mean is assigned to all missing responses.

- In the Survey on Income and Living Conditions, the method of class mean imputation is applied for calculating imputed rents. The classification variables, in order of importance, are as follows:
° *Type of housing*: The number of classes is limited to two: the household's accommodation is either a "one-family dwelling" or it is not ($i = 1,2$: where 2 is the number of classes)
° *Number of rooms*: It comprises four classes 1 or 2; 3; 4; 5 or more. ($j = 1,...,4$: where 4 is the number of classes)
° *Presence or otherwise of a place to sit outside:* Obviously, it comprises two classes (presence or otherwise of a place to sit outside. $k = 1,2$: where 2 is the number of classes)
° Assuming that each field contains at least 30 observations (if it does not, some observations will have to be grouped together), the imputed rent is the average of all respondent households in class $i\ j\ k$.

- Error localization by Fellegi-Holt based methods followed by regression (including ratio) imputation within classes, followed by adjustments in order to satisfy edit constraints.
- Used for continuous variables (for example: financial variables, number of employees)
- Random hot-deck within classes. To avoid inconsistencies and preserve correlations, related variables are imputed using the same donor. Used for continuous variables (for example: wages, total hours worked, number of holidays)
- Combined Edit and Imputation using Nearest-neighbour Imputation Methodology (NIM). This methodology will be used for both continuous (e.g. age) and discrete (e.g. marital status) variables. Application of this methodology is under development.
- Earnings: combination of historical imputation and nearest neighbour method (monthly earnings)
- Industry: combination of historical imputation and mean imputation (turnover and new orders)
- Income of an enterprise is the main variable in this survey. Some kinds of imputation are used for income. First of all – historical imputation. If there is no information on income from previous quarters, the average imputation is used. The average income for 1 employee in the stratum is calculated and then multiplied by the number of employees of the non respondent enterprise.
- In this survey imputation is also used for quantitative variables. Enterprises are asked about their financial indicators and the detailed information often contains mistakes. Imputation is used to correct them and also to compensate item non-response. Historical imputation and "nearest neighbor" methods are mainly used here.
- In Labour Force Survey imputation is used not very widely (because the response rate is quite high here, the respondents are interviewed), but for various variables (qualitative as well). Historical imputation, and "nearest neighbor" methods are used. In the nearest future data on wages and salaries from the State Social Insurance fund board should be available, so it will be used for imputation too.

3. How do you report and evaluate the effect of imputations on *the quality of estimates*? Please choose one of the above examples if you do not have opportunity to report about all examples. Possible alternatives are:
- we do not take care of it,
- we inform users about these operations including the advantages and disadvantages of imputations
- we estimate *imputation variance*, in addition to sampling variance.

Please tell something about your house style here and if you estimate *imputation variance* how do you do this?

In general, the items, that have been imputed, are marked. In some cases, we compute the estimates, before and after imputation, to evaluate the impact of the imputation. We don't inform users about the impact of imputation on the quality of estimates. We don't compute variance estimation due imputation. However, we carry out theoretical studies about the variance estimation taking account the imputed data under several imputation methods.

Based on our in house quality report system from which we are extracting a standard documentation for all statistical products *we inform the users about these operations including the advantages and disadvantages of imputation*. We do not calculate estimates for imputation variance.

We do not take care of it.

We inform users about these operations including the advantages and disadvantages of imputation

Imputation will never reproduce the true values (except in truly exceptional cases). The total error of the survey estimate is the sum of sampling error and the imputation error. Consequently, an overall variance is derived as the sum of a sampling variance and the imputation variance. (And next they present some formulae but it is not clear how much they will exploit these.)

Evaluation experiments are performed to asses the effects of imputation on estimates.

Imputation variance is not estimated on a regular basis.

Our house style is that we inform users at the end of the publication (so called "methodological explanations") about the level of unit- and item-nonresponse and how do we compensate for it (e.g. if imputation methods are used, which methods). In some cases, we calculate indices and their coefficients of variation before and after imputation is applied, but we do not publish the results.

At the moment these questions are subject to discussion regarding the release of quality reports by the NSI. So this question is simply about half a year to early. In the current practice the quality of estimates is not taken into account.

We don't calculate imputation variance and consider it as one of our "weak points" of the quality work which should be deleted. The users are informed about it by quality reports. The special questionnaire on quality of the main statistical indicators is being introduced now. The producers of statistics will be obliged to fill in it and to provide the quality information for the users. Questions on imputation are also included there.

The effects of imputation is studied mainly by comparing early estimates to final estimates. All known effects are corrected in the estimates. The imputation variance is not calculated

4. What is your opinion on the need for more international cooperation in this field? Thus how the international statistical society could develop appropriate and practicable imputation techniques and take into account the requirements of international comparisons and *the accuracy of estimates*? Other opinions about imputations?

The Eurostat quality report is a good tool to know the imputation process and to measure his impact in the final estimates. The United Nations Work Session on Data Editing is a useful forum to deal with these problems.

International cooperation is and was very useful in the field of imputation. As mentioned above we do not estimate imputation variance up to now but since the number of voluntary surveys with a high degree of non response is increasing the question of effects of imputation on the estimates is very important in the future.

Main topics on international research in the field of imputation could be:

- Imputation Variance for simple and complex estimators
- New Methods and their applicability for official statistics
  - Multiple Imputation
  - Neural Networks
- Evaluation of Imputation methods and publication strategy
  - How should users be informed
  - How to express the benefit of imputation

It is always a good idea to have common practices and comparable results. In certain cases though, imputation can happen using, for example, the experience of the NSI people rather than strict mathematical methods. Usually when you get the "real" results, empirical imputation is much more accurate. So, in terms of quality it is not always true that practicable imputation techniques are better.

It will be a great help

International cooperation is needed for the production of a suitable software so that the item non-response to be handled by some form of imputation.

We find international cooperation in this field useful if it is focused at the needs of NSI's, as our participation in the Euredit project shows.

We do not feel that any stronger international cooperation is necessary. There are a lot of new ideas that can be picked at some events (UN Work Session, International Workshop on Household Survey Nonresponse, various internet pages); usually the major problems are hidden in details when you're trying to implement the methods. Solution to these problems are usually country specific.

More cooperation would be appreciated by the FSO, not only with respect to an exchange of approaches and techniques, but also in a more fundamental manner, that takes into account the development of basic criteria for accuracy (like preservation of individual data, marginal distributions or the accuracy of estimates). Although the EUREDIT project already tackled these questions, there are still many question left open and some of the proposed criteria are at least questionable (e.g. with respect to criteria that emerged in the context of the theory of Multiple Imputation). This discussion should be closely related to quality issues.

International cooperation is very useful, especially keeping in mind the requirements of comparability of the surveys. We would appreciate and support if possible every effort and every activity in this field.

The main problem is to determine changes on time. Early recognition of new trends is vital

5. *Multiple imputation* (MI) is very popular in some research, but obviously not much used in NSI's. How is it in your NSI? Have you made any effort to exploit this method invented initially by D. Rubin and if yes, please tell more about your experience?

No. We have working more on resampling techniques.

We do not use MI up to now. However in a workshop concerning imputation we heard an interesting presentation how some estimators were generated using MI-methods applied on official micro data.

I think the main problem lies in the fact that the final product of an NSI is more or less something like final data file from which a publication is extracted. MI does not provide you with one data file with complete data sets. Of course theoretically it would be possible to generate a whole family of final data files after performing MI but this would yield to some data handling problems. A second point is that a publication of a NSI contains normally a lot of estimators. Performing MI for each one would increase the number of final data files dramatically.

I have never heard of it.

No, but I agree that is the best solution

Generally, the overall mean, class mean and regression imputations are methods, which do not work well for categorical and discrete variables. Consider now the estimation of the element standard deviation and distribution of the $y-\mathrm{var}iable$. Deterministic imputation methods (as class mean imputation) are not suitable for these purposes, since they cause attenuation in the standard deviation and they distort the shape of the distribution. By assigning the class mean to all the missing values in the class, the shape of distribution is clearly distorted with a series of spikes at the same class. The standard deviation of the distribution is attenuated because the imputed values reflect only the between-class and not the within-class variance.

The appeal for the *multiple imputation method* is that captures the within-class variance, and hence avoids the attenuation of the element standard deviation and the distortion of the distribution. Multiple imputation method replaces each missing value by a vector composed of $M \geq 2$ possible values. The major problem with the use of multiple imputations is the additional computer analysis needed, which increases as the number of M increases.

We have not exploited this method since most of our customers do not find the advantages to outweigh the burden of analyzing multiply imputed data sets.

We tested MI in the case of Quarterly Survey on Travel of Domestic Population, but we have not used it in practice. One of the problems are certainly users of micro data and their understanding.

MI is considered a challenging method for solving missing data problems in a coherent theoretical framework. In particular it shows, the necessity to improve imputation methods based on (parametric or semi-parametric) distribution approaches – instead of correcting data by clerical staff. The NSI started to evaluate the MI approach in some simulation studies (including modeling distributions by Artificial Neural Networks in order to derive MI imputation models).

Multiple imputation was tried to be used in Investments survey, but probably the variable (investment) wasn't the best choice for this method. We are going to try it for the more "stable" variables, like turnover in the survey of Inner trade.

We have not tried to apply MI to our data sets.

Some concluding points

Although this survey is not very representative, we can make some conclusions.

First, the imputation methods seem to be used much more for business and economic types of data. Consequently, most imputed variables are metric. This gives opportunities to make attempts with different types of imputation methods, thus including both model-donor and real donor methods. However, imputations are also used for some important categorical variables both in surveys and censuses, since there is a high need to fill in these data although some uncertainty may be occurred.

Second, there is a high range in imputation methods used. This survey does not show very well how some methods such as hot deck methods (*stochastic real-donor methods, random draw with replacement, all donor with an imputation cell as close to each other*) are really used. My general feeling is that the used methods are rather conservative focusing on improving such estimates as totals and averages, not so much taking care of distributions or associations. Typically imputations are also concerned some items, not often core variables, although these items have an effect on core variables. The imputed data seem to be mainly used at aggregate level, not so much for micro data analysis.

Third, most NSI's inform users on their imputation methods, but not all. On the other hand, the information is rarely very 'exact' so that users can know how much estimates are improved due to imputation and what is exactly the uncertainty due to imputation. It seems that NSI's are waiting for international standards, thus not much doing alone, except testing different solutions.

Fourth, most NSI's are willing to participate in international cooperation of this field, although maybe no new instruments are needed. This UNECE system, the International Nonresponse Workshop and Eurostat could be good forums for this purpose (see the above specific proposition on questions which could be appropriate in cooperation). Many NSI's also miss EU funded research projects such as Euredit, that gave rise for developing this kind of methodology. Unfortunately, there is not now any funding for statistical methodology research. Hopefully, the future will be better.

Fifth, I received interesting comments on multiple imputation. It is clear that no NSI (of these respondents) has really used this methodology but many have tested. I personally used MI early 1990's for disposable income but not exactly as Rubin proposed (see UNECE Statistical Journal 1992). Now, some NSI's seem to believe much to this technique, some have not heard anything about it. It was not clear for based on these answers whether all NSI's understand MI as an alternative technique for estimating imputation variance, or whether they consider it as a specific imputation method at the same time (MI thus can not be considered as a specific imputation method, it is an extension of a single imputation method always requiring adding stochasticity).
