WP. 8
ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Neuchâtel, Switzerland, 5-7 October 2009)

Topic (i): Automated editing and imputation and software applications

MULTIPLE IMPUTATION WITH STANDARD SOFTWARE: FIRST APPLICATION EXPERIENCES

Supporting Paper

Prepared by Katrin Schmidt, Federal Statistical Office of Germany

I. INTRODUCTION

1. Missing values are a common problem: How to deal with missing values is a question which has to be addressed in the course of any statistical data production. Missing values may be caused by item non-response, or they may occur at a certain stage of a data editing process: In a data editing process, some values are removed from the data set because they fail some edit rules. They should be replaced by values that will pass the edits. A wide variety of imputation methods has been discussed in the literature, see f.i. Schafer (2002) and Little and Rubin (2002). In this paper we focus on a special class of imputation methodology, i.e. multiple imputation (Rubin, 1978).

2. One of the central obligations of national statistical institutes is to compile data collected in a survey and to compute population (or sub-population) sums for the variables observed in the survey, like the number of unemployed academics or the turnover in a particular industry sector. Results are usually published in the form of statistical tables. Typically, data collection is based on random samples. In this case the sums are in fact statistical estimates. The quality of these estimates is usually assessed by variance estimation. For some data sets, more complex analyses are carried out for instance by external researchers in a data lab.

3. For obvious reasons, it is not always possible to actually observe the true value for any missing value in a data set (by, f.i. contacting the respondent a second time). Thus, imputations are usually estimates. This fact results in a well known problem for any single imputation method (e.g. a method which imputes just one (single) value to replace a missing value): After completing the data by imputation, data analysis proceeds as if there had been no missing values. The additional uncertainty in the population estimates caused by the fact that a part of the observations are actually only estimates of these observations is then not taken into account in the analyses. Consequently, the standard errors provided by variance estimation are underestimated, and thus the size of the confidence intervals for the estimates of the (sub-) population totals. A multiple imputation approach on the other hand provides a variance estimator that takes into account the additional uncertainty due to imputations.

4. In this paper we report on first experiences and test results obtained when trying to apply the multiple imputation facility of the standard software IVEware (Ragunathan, Solenberger and Van Hoewyk 2002) to a business survey data set. These experiments are a first step in a much wider research context which in the end should answer questions regarding practical usability of multiple imputation techniques in the German statistical system.

5. In section II we briefly sketch the methodology implemented in IVEware. After describing the test data and the test scenario in III.A, section III.B introduces some indicators we use to evaluate our test results. The pre-processing which is necessary to come to a suitable input data set for IVEware is explained in section III.C. Section IV presents some test results and gives an outlook to future work. The paper ends with a summary section in V.

II. METHODOLOGICAL BACKGROUND

6. IVEware[1] performs single or multiple imputations of missing values using the sequential regression imputation method described in Raghunathan, Lepkowski, Van Hoewyk and Solenberger, (2001). In the following we briefly summarize the main concepts of this method. While IVEware produces its own test diagnostics to evaluate the results of an application, we introduce some additional quality indicators explained in III.B.

7. The general idea of multiple imputation is to replace a missing value in a data set not by imputing just one single value to fill the gap, but by generating m ($m > 0$) values for imputation. In this way, we end up with m completed data sets. Each of these datasets will be analyzed separately (f.i. to retrieve population estimates) and the m results will be combined finally, using some simple formulas. If, for example, the goal is to estimate a population total Q, the following formulas can be used:

$\hat{Q}_{MI}$ is the multiple imputation estimate for Q: $\hat{Q}_{MI} := \bar{\hat{Q}} = \frac{1}{m}\sum_{d=1}^{m}\hat{Q}^{(d)}$,

where $\hat{Q}^{(d)}$ refers to the estimate of Q retrieved from separate analysis of the d-th completed data set.

T_{MI} is the combined variance estimate for the estimate $\hat{Q}_{MI}$: $T_{MI} := \bar{S}_{MI} + \left(1+\frac{1}{m}\right)B_{MI}$ ("total variance"),

where $\bar{S}_{MI} := \frac{1}{m}\sum_{d=1}^{m}S^{(d)}$ ("within variance") , $B_{MI} := \frac{1}{m-1}\sum_{d=1}^{m}\left(\hat{Q}^{(d)} - \hat{Q}_{MI}\right)^2$ ("between variance") and $S^{(d)}$ refers to the variance estimate for the estimate $\hat{Q}^{(d)}$, estimated in a separate analysis of the d-th completed data set.

8. The sequential regression imputation method of Raghunathan et al. (2001) as implemented in IVEware and as accurately described in Tempelmann (2006) computes imputations using a Bayesian regression approach, iterating a sequence of regressions. At each step of, say, t iterations, imputations are drawn from the posterior predictive distribution that is given by the regression model and an uninformative prior. The number of regressions in a sequence corresponds to the number of variables of the data set with missing values to be filled by imputed values. For each of these variables a regression is fitted, where the predictors are certain auxiliary variables and the other variables with (initially) missing values. The parameters of the regression are drawn from a posterior distribution, which is estimated using observed and previously imputed values. These parameters are then used to draw new observations to impute for the missing values of the dependent variable from its conditional density, i.e. according to the regression model, given the observed and previously imputed values of the current independent variables.

9. When modelling regressions the IVEware package takes into account variable specific properties, such as semi-continuity. Semi-continuous variables take on a single discrete value (e.g. zero) with positive probability, but are continuously distributed otherwise. For variables of this type, the software uses a two stage procedure. In the first stage the zero/non-zero status of the variable is determined by using a logistic regression, where the dependent variable takes the values one and zero. A value of zero (resp. one) of the dependent variable in the logistic regression implies a value of zero (resp. non-zero) of the variable which is to imputed. In the second stage a truncated normal regression model is fitted restricted to records with non-zero status of the variable.

[1] IVEware=Imputation and Variance Estimation Software

Moreover, IVEware can also handle linear restrictions on variables (like e.g. resulting from edit restrictions) as well as model interactions between variables on account of nonlinear relationships. Apart from continuous and semi-continuous variables the package can also deal with other types of variables, such as categorical or count variables.

III. THE TEST APPLICATION

10. We begin this section by describing our test data set and the test scenario (III.A) . In III.B we introduce quality indicators to evaluate test results and explain some data preparation steps that have to be taken in advance of running the IVEware package in III.C and D.

A. Test data

11. We are examining an implementation of multiple imputation with data of the structural business survey in the sector wholesale and wholesale on a fee or contract basis. These data are collected annually by a stratified random sample. We only got test data from the 2004 survey, as raw data prior to editing and imputation are only available for this period of the survey. Our test data set consists of 9,326 records and 25 variables. The variables are listed in table 1.

12. Some variables of the raw data set show an enormous amount of missing values. For the variable "subventions" (no. 2 in table 1), for example, 90 per cent of the data are missing. In most cases the missing values are in fact zeroes as in the raw data no zeros are present. For this reason it is imperative to distinguish between "truly" missing values and "not truly" missing values by putting zeroes in the gaps where the values are not truly missing before doing a multiple imputation. Otherwise IVEware will impute values greater than zero, leading to violations of edit constraints. When making this distinction we have to be very careful, because it has a strong impact on the quality of the data to be published. The decision whether a value is "truly" missing or not is made by formulating additional edit constraints. When certain relations between variables hold, it can be deduced that a missing value is in fact a zero. After having filled zeroes into those gaps we call the raw data "*pre-processed raw data*".

13. It is these zeroes that are special in this statistic. Trade is very flexible; the products and activities offered can vary from one data collection to the next, and a company can both trade and at the same time be a wholesale company on a fee or contract basis. Furthermore, the wholesale on a fee or contract basis companies play a special role in a way: they have positive, non-zero values for turnover but no expenses nor inventories. Therefore, these zeroes are structural zeroes for one part of the records and non-structural zeroes for another, and may consequently occur in large numbers.

14. Since our pre-processed raw data has not yet undergone a complete editing process, there are erroneous records in the pre-processed raw data. Comparing the pre-processed raw data to the edited data of the annual survey of 2004 we identified the cases with a difference between the value in the edited data set used as data basis for the published estimates and the value in the pre-processed raw data, e.g. the value in the corresponding entries had been corrected during the editing process. So we consider these entries as erroneous in the pre-processed raw data set. The entries thus identified are then deleted.

15. We do not have complete "original" (i.e. true) data of our data set. Therefore it is not possible to make statements about the quality of the multiple imputations if the method is applied to the entire data set. For this reason only the set of records without errors was used for our testing (henceforth referred to as "*correc*t *data set*"). These are the data-records that do not show any missing values after comparison of the pre-processed data set with the published data set. 5,851 records are without errors (obtained through a pre-processing with the complete set of edit rules as described in III.D). We generated missing values in this subset of records by a simple random sampling. We set the selection probability equal to the fraction of missing values in the full pre-processed data set consisting of 9,326 records after deleting the entries which are considered as erroneous.
This method of generating missing values results in missing data which is missing completely at random (MCAR), i.e. the simplest situation of all.
The resulting amount of the missing values in the individual variables in the 5,851 records is shown in table 1. In the following we refer to this data set as "*test data set*".

Table 1: *Description of the test data set.*

VARIABLE	DESCRIPTION	MISSING VALUES	ZEROS	DISTRIBUTION
1.Turnover	turnover	48	0	normal
2.Turnover_Sub	turnover of subsidies	4	5623	mixed
3. Inv_Trade_Begin	inventories of trade goods at the beginning	295	1064	mixed
4. Inv_Trade_End	inventories of trade goods at the end	408	1007	mixed
5.Inv_Raw_Begin	inventories of raw materials at the beginning	313	4488	mixed
6.Inv_Raw_End	inventories of raw materials at the beginning	320	4472	mixed
7.Exp_Trade	expenditures for trade goods	264	288	mixed
8.Exp_Raw	expenditures for raw materials	227	4140	mixed
9.Exp_Wages	expenditures for wages	108	392	mixed
10.Exp_SSC	expenditures for social security contributions	176	392	mixed
11.Exp_Rent	expenditures for rent	33	968	mixed
12.Exp_Tax	expenditures for tax	507	0	normal
13.Exp_Other	other expenditures	830	0	normal
14.Invest_ Real_Est	investments in real estates	2	5596	mixed
15.Invest_Ex_Build	investments in existing buildings	2	5511	mixed
16.Invest_Er_Build	investments in erected buildings	4	5310	mixed
17.Invest_Mach	investments in machinery	11	2388	mixed
18.Invest_Tang_Ass	sales of tangible assets	8	3777	mixed
19.Invest_Leas_Tang_Ass	value of sold tangible assets by financial leasing	1	5399	mixed
20.Pers_Part	persons in part-time	196	1252	?
21.Pers_Owner	persons owner	148	3360	poisson
22.Pers_Emp_Lab	persons employees and labourer	267	372	?
23.Pers_Other	other persons	36	5532	?
24.Pers_Female	persons female	92	512	?
25.Pers_Male	persons male	115	113	?

B. Quality Indicators

16. We are particularly interested in the difference between population totals estimated on the basis of the completely observed random sample and the corresponding population totals estimated on the basis of the test data after multiple imputation. A comparison of their confidence intervals is especially interesting. To this end we compare the estimates of the “correct” data to the estimates of the test data. We define quality indicators as follows:

17. Let

- ***Rel_Conf_L***:= *„Lower bound of the confidence interval of the estimate of the multiply imputed data relative to the lower bound of the confidence interval of the estimate of the completely observed data“.*
  Rel_Conf_U:= *„Upper bound of the confidence interval of the estimate of the completely observed data relative to the upper bound of the confidence interval of the estimate of the multiply imputed data “.*

18. Since our test data are non-negative, the indicators ***Rel_Conf_L*** and ***Rel_Conf_U*** should be ideally close to 1 and less or equal to 1. In this case the confidence interval of the estimate based on the multiply imputed data covers the confidence interval of the estimate based on the completely observed data and the loss of precision is small.

19. Another interesting indicator can be:

- The relative difference between the estimate computed of the multiply imputed data set to the estimate computed of the correct data set on basis of the estimate computed of the correct data set:

$$Rel\_Diff := \frac{\hat{Q}_{org} - \hat{Q}_{MI}}{\hat{Q}_{org}} * 100\%,$$

  with denotations from section II.

20. Furthermore, we compare the standard deviation of the estimate resulting from the multiply imputed data ($\sqrt{T}$) to the standard deviation of the estimate computed for the correct data ($\sqrt{S_{org}}$), using the following indicator

$$Rel\_Std := \frac{\sqrt{T}}{\sqrt{S_{org}}},$$

where S_{org} denotes the common variance estimate from the correct data set corresponding to the sampling design.

Rel_Std should be greater than or equal to 1. Values less than 1 imply that the imputed values are near the means of the observed test data, i.e. the multiply imputed data exhibit less variation than the correct data. Hence, values less than 1 indicate that the imputation model does not work well.

C. Transformation and initial choice of variables

21. Before application of IVEware we must transform our variables to f.i. approximate normality. By means of Box-Cox transformation (see e.g. Draper and Smith (1998)) we were able to transform most of the variables successfully to one of the distributions that can be modelled by IVEware as shown in table 1. For some variables, however, none of those distributions seem to work. Here we have to do some further investigation. For now, we only take into consideration variables 1 to 19 from table 1 which were successfully transformed.

22. In our first trial of a multiple imputation application we do not model interactions of variables. Hence, each sequence of regressions consists of simple linear and logistic regressions. As mentioned in II, the logistic regression is part of a two-stage approach for the imputation of semi-continuous variables. The two-stage approach consists of a logistic regression which is followed by a linear regression, depending on the outcome of the logistic regression. Also, in this first trial we do not set restrictions to variables nor do we specify bounds to imputed values. Accordingly, our settings in IVEware are as follows: choice of the distribution of the variables, definition of the number of iterations of the algorithm

and the number of imputations, and specification of the seed of the algorithm. We created 36 versions by multiple imputation with the test data: for each of three different seeds we generated 10 imputations by $t \in \{1,2,3,4,5,6,8,10,15,20,30,40\}$ iterations.

D. Separation Problems – Extension of the Pre-Processing

23. A first application of IVE-Ware to the test data set (obtained through a pre-processing with an incomplete set of edit rules) with variables 1 to 19 failed. The warning messages of IVE-Ware described difficulties of the algorithm caused by multicollinearity of the variables. Looking into the data, we found that the problems were caused by the phenomenon of (quasi-) separation and resulted in a non-existence of finite estimators of the logistic regressions. As mentioned in II, imputations of semi-continuous variables are generated by a two-stage approach using a logistic regression.
Separation in logistic regression frequently occurs when the binary dependent variable can be perfectly separated by a single independent variable or by a non-trivial linear combination of the independent variables (Albert and Anderson (1984)). Quasi-separation occurs when the binary dependent variable can be almost perfectly separated by a single independent variable or by a non-trivial linear combination of the independent variables.
In our case the quasi-separation of the binary dependent variable in the two-stage approach of a semi-continuous variable resulted from a lack of edit rules in our pre-processing of the raw data. We had created too few edit rules. The correct data set (obtained through a pre-processing with an incomplete set of edit rules) reveals a quasi-separation e.g. of the variables Inv_Trade_Begin (no. 2 in table 1) and Inv_Trade_End (no. 3 in table 1). However, the data set which is the basis for the published estimates does not exhibit a separation of variables. In this data set a company without inventories of trade goods at the beginning (no. 2 in table 1) of the survey period can accumulate inventories during the survey period. On the other hand in this data set it also happens that a company having inventories at the beginning (no. 3 in table 1) of the survey period can have no inventories at the end of the survey period. Furthermore there are no non-zero values which separate the data set. However in our correct data set occurs a separation of the data as we did not set a zero in any field of the variable Inv_Trade_End where the variable Inv_Trade_Begin exhibits a value greater than zero. Thus, there is a quasi-separation since a value zero of the variable Inv_Trade_End implies a value zero of the variable Inv_Trade_Begin and a value greater than zero of the variable Inv_Trade_End does not imply a value zero or greater than zero. Tables 2 and 3 contain the combination of zero/nonzero values of both variables.
Therefore we created additional (sensible) edit rules for the pre-processing of the raw data set which eliminate the separation of variables. As a consequence, the resulting correct data set (5,851 records) is larger than the previous correct data set (5,764 records) as now more records are correct and complete. These problems, which occurred because we did not create enough edit rules, makes us aware of the fact that the necessary pre-processing of the raw data is quite essential and has to be done very cautiously.

Tabelle 2: *Combinations of zero and non-zero values of the variables Inv_Trade_Begin and Inv_Trade_End in the data set which is the basis for the published estimates. "+" means "this combination exits in this data set" and "-" means "this combination does not exist in this data set"*

	Inv_Trade_Begin = 0	**Inv_Trade_Begin > 1**
Inv_Trade_End = 0	+	+
Inv_Trade_End > 0	+	+

Tabelle 3: *Combinations of zero and non-zero values of the variables Inv_Trade_Begin and Inv_Trade_End in the correct data set (obtained by pre-processing with an incomplete set of edit rules). "+" means "this combination exits in this data set" and "-" means "this combination does not exist in this data set"*

	Inv_Trade_Begin = 0	**Inv_Trade_Begin > 1**
Inv_Trade_End = 0	+	-
Inv_Trade_End > 0	+	+

IV. TEST RESULTS

24. After the extension of the pre-processing described above we were able to perform the 36 variants of multiple imputations without receiving any warning messages. Nevertheless, our indicators show instabilities of the algorithm by enormous high or small imputed values. Fortunately, these abnormalities are present for just a few variants. Currently we are unable to identify the cause of this problem but we assume is has to do with variables 14 to 19. Since the number of missing values in these variables is small, we ignore them for now.

25. After omitting variables 14 to 19, the imputation variants finish without warning messages and the indicators do not reveal any too strong anomalies. An extract of the results is contained in the appendix in tables 4 to 7. Looking at our indicators the results for many variables are quite convincing. The outcomes of the multiple imputation estimates of the sums and their confidence intervals of the variables 1, 2, 3, 4, 7, 9 in table 1 are close to their correct estimates. However, the results for the variables 5, 6, 8, 10, 11 in table 1 are less satisfactory. The results for the variable Exp_Raw (no. 8 in table 1) are particularly suspicious as the indicator Rel_Std is less than one for most imputation variants and exhibits one of the highest values of all variables. That is, with few exceptions, its standard deviation estimator of the multiply imputed data set is smaller than that of the correct data set. This may indicate defects of the imputation model for this variable. Furthermore, the quality indicators introduced in III.B of the variable 5, 6, 10, 11 reveal that we get different results of the survey for these variables when missing data is present: the quality indicator Rel_Dev introduced in III.B (except variable no.6) show that the relative deviation is quite large. Also the quality indicators Rel_Conf_L and Rel_Conf_U exhibit that the confidence intervals of the multiply imputed data set do not include the confidence intervals of the correct data set.
Possible causes of the less satisfactory results for the variables 5, 6, 10, 11 are manifold. Perhaps the results can be improved by modelling the sampling design of the survey (a stratified sampling design was designed). Additionally, we have not considered non-linear relationships of variables yet. Therefore the results might be enhanced by modelling non-linear relationships of variables.

V. OUTLOOK: FUTURE WORK

26. In order to figure out the reasons of the problems mentioned above, we will have to continue the testing, addressing in particular the following questions:

1. Can the quality of the multiply imputed data be improved by adjusting the imputation model to account for the sampling design? Perhaps by including dummy variables?
2. Can the quality of the multiply imputed data be improved by modelling interactions of the variables?
3. Are outliers (jointly) responsible for the problems?
4. Are strongly correlated variables (jointly) responsible for the problems?

27. Furthermore, we have to examine if and how we can include the skipped variables into our application. In addition to that we will check if our imputations are consistent with the edit rules. Since we have not set linear restrictions yet, we expect that some records will fail the edits.
Besides, we have to generate missing data in our correct data set by different sampling designs. Then we have to run our tests with the resulting test data sets to compare the multiple imputation results for different missingness mechanisms.

V. SUMMARY

28. In this paper we describe a first experiment to apply the standard software IVEware for multiple imputation to a real data set from an economic survey. After solving initial difficulties by omitting some variables and an extended pre-processing the application provided promising results. For many variables the overlap between confidence intervals for population estimates derived from the correct data set as compared to the imputed data set is nearly perfect.
However, as mentioned in section V. before putting this into practice, we will have to gather much more experience with the methodology and to come to a better understanding of the problems that may occur.

It should then be possible to find ways to impute also variables that had to be omitted previously. Using more elaborate modelling etc. it may also be possible to improve the performance of the method for some of the variables.

References

Albert, A., Anderson, J. (1984): "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models", Biometrika, Vol. 71 pp. 1-10

Draper, N., Smith, H, (1998): "Applied Regression Analysis", Wiley Series in Probability and Statistics

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

Raghunathan, T., Lepkowski, J., van Hoewyk, J., and Soenberger, P. (2001): "A multivariate technique for multiply imputing missing values using a sequence of regression models", Survey Methodology, 27, pp. 85-95

Raghunathan, T.E., Solenberger, P., Van Hoewyk, J. (2002): "IVEware: Imputation and Variance Estimation Software User Guide", http://www.isr.umich.edu/src/smp/ive/

Rubin, D.B. (1978): "Multiple Imputation in Sample Surveys - a Phenomenological Baysian Approach to Nonresponse", American Statistical Association Proceedings of the Section on Survey Research Methods, pp.20-40

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

Schafer, J., Graham, J. (2002): "Missing Data: Our View of the State of the Art", Psychological Methods, Vol. 7, No. 2, pp. 147-177

Tempelmann (2006): "Imputation of Economic Data Subject to Multiple Linear Restrictions Using a Sequential Regression Approach", Work Session on Statistical Data Editing 2006

Appendix

Table 4*: (Rel_Conf_L) Lower bound of the confidence interval of the estimate of the imputed data relative to the lower bound of the confidence interval of the estimate of the correct data. X_Y means seed "X" and "Y" iterations.*

Variable\Iteration	1_1	1_2	1_3	1_30	1_40	3_1	3_2	3_3	3_30	3_40
Turnover	1,00	1,00	1,00	1,00	1,00	1,01	1,00	1,00	1,00	1,00
Turnover_Sub	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Inv_Trade_Begin	1,01	1,00	1,00	1,00	1,00	1,01	1,00	1,00	1,00	1,00
Inv_Trade_End	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Inv_Raw_Begin	1,03	1,02	1,05	1,02	1,06	0,64	1,05	1,05	1,06	1,04
Inv_Raw_End	1,01	1,00	1,01	1,01	1,01	1,01	1,01	1,00	1,00	1,01
Exp_Trade	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Exp_Raw	1,00	0,98	0,98	0,98	0,98	0,96	0,98	0,98	0,77	0,98
Exp_Wages	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Exp_SSC	0,99	0,99	0,99	0,99	0,99	0,97	0,97	0,99	0,97	0,99
Exp_Rent	1,02	1,00	1,00	1,00	1,00	1,01	1,00	1,00	1,00	1,00
Exp_Tax	1,02	0,94	1,01	1,02	1,01	1,00	1,01	1,01	1,01	1,02
Exp_Other	1,02	1,02	1,03	1,03	1,03	1,03	1,03	1,03	1,02	1,02

Table 5: *(Rel_Conf_U) Upper bound of the confidence interval of the estimate of the correct data relative to the upper bound of the confidence interval of the estimate of the imputed data. X_Y means seed "X" and "Y" iterations.*

Variable\Iteration	1_1	1_2	1_3	1_30	1_40	3_1	3_2	3_3	3_30	3_40
Turnover	0,99	1,00	1,00	1,00	1,00	0,99	1,00	1,00	1,00	1,00
Turnover_Sub	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Inv_Trade_Begin	0,99	1,00	1,00	0,99	1,00	0,99	1,00	1,00	1,00	1,00
Inv_Trade_End	0,96	1,00	1,00	1,00	1,00	0,99	1,00	1,00	1,00	1,00
Inv_Raw_Begin	0,90	0,98	0,95	0,98	0,95	0,63	0,96	0,96	0,94	0,97
Inv_Raw_End	0,99	1,00	0,99	0,99	0,99	0,99	0,99	1,00	1,00	0,99
Exp_Trade	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Exp_Raw	1,00	1,04	1,02	1,02	1,03	1,05	1,03	1,04	0,83	1,04
Exp_Wages	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Exp_SSC	1,01	1,00	1,01	1,01	1,00	0,98	0,97	1,01	0,95	1,00
Exp_Rent	0,98	1,00	1,00	1,00	1,00	0,96	1,00	1,00	1,00	1,00
Exp_Tax	0,96	0,84	0,98	0,93	0,98	0,99	0,97	0,97	0,97	0,97
Exp_Other	0,96	0,97	0,93	0,94	0,96	0,95	0,97	0,96	0,97	0,94

Table 6*: Relative deviation of the estimate of the correct data to the estimate of the imputedt data in %. X_Y means seed "X" and "Y" iterations.*

Variable\Iteration	1_1	1_2	1_3	1_30	1_40	3_1	3_2	3_3	3_30	3_40
Turnover	-0,53	-0,11	-0,02	-0,08	-0,11	-0,55	-0,04	-0,03	-0,04	0,02
Turnover_Sub	0,00	-0,01	0,00	0,00	-0,06	0,00	-0,01	-0,01	0,00	-0,01
Inv_Trade_Begin	-0,84	-0,27	-0,21	-0,58	-0,01	-0,73	-0,02	-0,25	-0,36	-0,41
Inv_Trade_End	-1,84	-0,10	0,16	-0,03	0,00	-0,97	0,03	-0,14	-0,01	-0,06
Inv_Raw_Begin	-8,23	-2,12	-5,34	-1,90	-5,29	-24,51	-4,69	-4,26	-5,87	-3,33
Inv_Raw_End	-0,95	0,27	-0,88	-0,55	-1,09	-1,16	-0,84	-0,22	0,25	-0,73
Exp_Trade	-0,18	-0,10	-0,06	-0,27	-0,18	-0,38	-0,17	-0,27	-0,11	0,00
Exp_Raw	-0,16	3,01	2,09	1,62	2,63	4,41	2,70	3,36	-3,50	2,83
Exp_Wages	-0,02	-0,03	-0,05	0,02	-0,19	0,04	-0,07	-0,01	0,05	-0,11
Exp_SSC	0,93	0,47	0,86	0,88	0,50	-0,07	-0,62	0,70	-1,43	0,62
Exp_Rent	-1,72	-0,12	-0,22	-0,25	-0,08	-2,97	-0,18	-0,21	-0,26	-0,26
Exp_Tax	-3,13	-7,71	-1,79	-4,95	-2,01	-0,55	-2,54	-2,37	-2,03	-2,68
Exp_Other	-3,22	-2,21	-5,31	-4,77	-3,97	-4,16	-2,71	-3,73	-2,52	-4,20

Table 7: *Rel_Std. X_Y means seed "X" and "Y" iterations.*

Variable\Iteration	1_1	1_2	1_3	1_30	1_40	3_1	3_2	3_3	3_30	3_40
Turnover	1,01	1,00	1,00	1,00	1,00	1,01	1,00	1,00	1,00	1,00
Turnover_Sub	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Inv_Trade_Begin	1,02	1,01	1,01	1,02	1,01	1,02	1,01	1,01	1,01	1,01
Inv_Trade_End	1,31	1,01	1,01	1,00	1,01	1,08	1,00	1,01	1,01	1,00
Inv_Raw_Begin	1,21	1,03	1,05	1,01	1,04	2,75	1,04	1,02	1,06	1,02
Inv_Raw_End	1,01	1,00	1,01	1,00	1,01	1,01	1,01	1,00	1,00	1,01
Exp_Trade	1,00	1,00	1,00	1,01	1,00	1,01	1,00	1,01	1,00	1,00
Exp_Raw	1,00	0,94	0,97	0,99	0,97	0,92	0,95	0,93	2,01	0,94
Exp_Wages	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Exp_SSC	0,96	1,03	0,99	1,03	1,01	1,29	1,39	0,98	1,47	1,06
Exp_Rent	1,01	1,00	1,00	1,00	1,00	1,13	1,00	1,00	1,00	1,00
Exp_Tax	1,11	**2,26**	1,07	1,28	1,07	1,04	1,13	1,10	1,10	1,05
Exp_Other	1,16	1,09	1,27	1,22	1,12	1,14	1,04	1,15	1,08	1,32