

Generated Asset Variables in the Survey of Health, Aging and Retirement in Europe¹

Dimitrios Christelis²
University of Salerno and CSEF

Tullio Jappelli
University of Salerno, CSEF and CEPR

Mario Padula
University of Salerno and CSEF

April 25, 2005

Table of Contents

1. Introduction.....	2
2. Definitions.....	2
2.1 Amounts.....	2
2.2 Flags.....	4
2.3 Top-coding	6
3. Variable Names.....	7
4. Calculation of Purchasing Power Parities.....	11
5. Imputation.....	12
5.1 Imputation of Ownership Variables	12
5.2 Imputation of Amount Variables.....	13
5.3 Imputation of Indirect Stockholding.....	15
5.4 Multiple Imputation.....	15
6. Brief Description of the Program that generates the Asset Variables.....	17
7. References.....	18

¹ This memo refers only to Release 1 data

² Corresponding author, e-mail: cdimitri[at]unisa.it We are grateful to Stephanie Stuck for her assistance and suggestions

1. Introduction

This note describes the construction of the wealth-related variables in SHARE, namely their definition and naming, the purchasing power adjustment, the imputation procedures used, and the structure of the program that performs these calculations.

2. Definitions

2.1 Amounts

First, the following individual-level magnitudes are generated (the question names to which they correspond are in parentheses):

- i) Value of the primary residence (HO024_)
- ii) Value of the mortgage (HO015_)
- iii) Value of other real estate (HO027_)
- iv) Value of bank accounts (AS003_). When the value of this variable is initially negative, it is set to zero and the negative part is added to the financial liabilities
- v) Value of government and corporate bond holdings (AS007_)
- vi) Value of stock holdings (AS011_)
- vii) Value of mutual fund holdings (AS017_)
- viii) Value of individual retirement accounts (AS021_, AS024_)
- ix) Value of the contractual savings for housing (AS027_)
- x) Value of life insurance policies (AS030_)
- xi) Value of owned business, including the non-owned part of it (AS042_)
- xii) Owned share of own business (AS044_)
- xiii) Value of owned cars (AS051_)
- xiv) Value of financial liabilities (AS055_) plus the negative bank account balances.

By multiplying xi) by xii) above one obtains:

xv) Value of owned share of own business.

In addition, we impute the value of risky assets, which we define to be direct stock holdings, and the percentage of holdings in mutual funds and individual retirement accounts that are invested in stocks. We cannot directly observe the latter two quantities. We have however questions for both mutual funds (AS019_) and individual retirement accounts (AS023_, AS026_), which give information on whether the amount invested is mostly in stocks, roughly equally in stocks and bonds or mostly in bonds. We impute respectively to these three possible answers the following percentages of investment in stocks: 75%, 50% and 25%. Using this imputation we construct:

xvi) Value of holdings of risky financial assets

At a second stage, the individual-level variables i, ii, iii, iv, v, vi, vii, viii, ix, x, xiii, xiv, xv and xvi defined above are summed over all household members in order to generate the corresponding household-level variables. In addition we generate the following household-level aggregates:

xvii) *Real assets* are defined as the sum of the value of the primary residence net of the mortgage, the value of other real estate, the owned share of own business and the owned cars.

xviii) *Gross financial assets* are equal to the sum of the values of bank accounts, government and corporate bonds, stocks, mutual funds, individual retirement accounts, contractual savings for housing and life insurance policies owned by the household.

xix) *Net financial assets* are equal to gross financial assets minus financial liabilities.

xx) *Risky financial assets* are equal to sum of direct stockholding and the imputed share of mutual funds and individual retirement accounts invested in stocks.

xxi) *Net worth* is equal to the sum of real and net financial assets

There are some deviations in the definitions of some assets across countries. Individual retirement accounts are not included for the Netherlands, whereas in Austria a different kind of retirement accounts is reported, called *Praemiengefoerderte Zukunftsvorsorge*, which are state-run.

In addition, there are two kinds of life insurance in France:

- a) Life insurance that is an annuity paid as long as the policy holder remains alive, which is counted as part of the individual retirement accounts.
- b) Life insurance that is paid to survivors in case of death of the policy holder, which is counted as part of life insurance holdings

2.2 Flags

In addition to generating the variables for the wealth-related items, we need to generate also their corresponding flag variable, which contains information about how the amount variables were constructed. For individual-level variables the flag variable takes the following values:

- 1 - *Continuous answer*: the respondent answered with a positive or negative value to the amount question, and there was no need to amend her answer in any respect.
- 2 - *Complete Bracket*: the respondent did not want or did not know how to answer the amount question, but then entered into the unfolding bracket procedure and successfully completed it.
- 3 - *Incomplete Bracket*: the respondent did not want or did not know how to answer the amount question, entered into the unfolding bracket procedure but did not complete it.
- 5 - *Refusal to start the bracket sequence*: the respondent did not want or did not know how to answer the amount question, and again refused or did not know how to answer the first unfolding bracket question.
- 6 - *No ownership*: the respondent does not own the item.

- 7 - *Refusal/Don't know on ownership question*: the respondent refused or did not know how to answer the question on ownership that precedes the amount question for each item.
- 9 - *Is not a financial respondent*: the respondent is not the designated financial respondent for the household and does not report any amount for the item.
- 10 - *Negative values, 0s, implausibly low positive values, wrong currency answers, very high outliers*: this broad category includes cases for which it was decided that the values were so implausible as to be a result of some mistake or an alternative form of refusal to answer the question. For these cases we used imputation to fill in the values.

The flag variable takes the following values for household-level variables:

- 0 - *Doesn't own the item*: no household member owns the item in question
- 1 - *No imputation*: there has been no imputation done for any household member for the item in question
- 5 - *Some imputations*: there has been imputation for at least one household member for the item in question or, in case of a composite item, for at least one of its constituent parts
- 9 - *No housing/asset/liability section respondent*: there is no respondent for the particular household in the housing/asset/liability section of the survey, but imputation is still performed.

Some clarifications are needed for the last value of the flag variable for individuals. We treated negative values as implausible, with the exception of bank accounts and the value of own business. The balance of the former can be negative because of overdrafts for example, and the latter's value can be negative when the assets of the business are less than its liabilities.

There are some cases for which the amount is stated to be zero, while the ownership variable is positive. This might be an indication of refusal to answer the amount question,

without going into the unfolding brackets procedure. We consider these cases to be missing and we impute them.

For countries that have adopted the euro as their currency (i.e. Austria, Belgium, France, Germany, Greece, Italy, Netherlands, Spain), the respondent can give an answer to an amount question either in euros or in pre-euro currency. Unfortunately, some answers in pre-euro currency are entered by mistake as an answer in euros and vice-versa. Given that the exchange rate of the old currency to the euro is always greater than one, this mistake can be detected only for countries for which the euro conversion exchange rate is very high, namely Italy (exchange rate equal to 1936.27), Greece (340.75), Spain (166.39), and possibly Austria (13.76), Sweden (9.18), Denmark (7.44) and France (6.56), and for answers with unusually high values in euros or for unusually low values in local currency. In determining whether an answer is entered in the wrong currency column we also take into account whether the respondent has answered other questions in pre-euro currency. When the answer is deemed to be entered in the wrong currency, we divide or multiply by the exchange rate.

Finally, after correcting for a wrong currency entry, we are still left with some implausibly high outliers. The thresholds above which a value is considered to be a high outlier are: 15,000,000 for the primary residence, the mortgage, other real estate, bank accounts, stocks and mutual funds, 10,000,000 for bonds and financial liabilities, 5,000,000 for individual retirement accounts, contractual savings, life insurance holdings and cars and 50,000,000 for the value of an own business. We set the values above those thresholds to missing and impute them, conditional on being on the highest bracket.

2.3 Top-coding

Swedish data has been top-coded (i.e. values have been modified to equal a maximum level if they originally supersede it) due to legal constraints, according to the Swedish Secrecy Act. The variables that are subject to top-coding are:

- 1) ho024_ (value of the primary residence), with a top-coding threshold of 9,000,000 SEK or 1,000,000 euros

- 2) ho027_ (value of other real estate), with a top-coding threshold of 25,000,000 SEK or 2,750,000 euros
- 3) hnetwv_e (household net worth in euros), with a top-coding threshold of 15,000,00 euros.

Variables hnetwv_p (household net worth in euros with adjustment for purchasing power parity), hrav_e and hrav_p (household value of real assets, in euros and purchasing power parity-adjusted euros respectively) are calculated obeying the aforementioned top-coding constraints.

3. Variable Names

In the table below one can find the names of the variables that are generated by the asset working group. The suffix _e signifies that the answer is in euros (after conversion from an original non-euro currency answer where applicable) while the suffix _p denotes a conversion of the amount in euros to an amount adjusted to reflect the differences in the price levels between countries. This adjustment is made using purchasing power parities (PPP) provided by the Organization for Economic Co-operation and Development (OECD) and is described in section 4 below. The suffix _f denotes the flag variable corresponding to a particular item and the prefix h signifies that the variable is computed at the household level.

a. Dataset indexing variable

implicat:	Indexes each dataset generated by multiple imputation (see Section 5.4 below)
------------------	---

b. Individual-level variables

homev_e:	Value of the house, not PPP-adjusted (euros)
mortv_e:	Value of the mortgage, not PPP-adjusted (euros)
oresv_e:	Value of other real estate, not PPP-adjusted (euros)
baccv_e:	Value of bank accounts, not PPP-adjusted (euros)
bondv_e:	Value of bonds, not PPP-adjusted (euros)

stocv_e:	Value of stocks, not PPP-adjusted (euros)
mutfv_e:	Value of mutual funds, not PPP-adjusted (euros)
irav_e:	Value of individual retirement accounts, not PPP-adjusted (euros)
contv_e:	Value of contractual savings for housing, not PPP-adjusted (euros)
linsv_e:	Value of life insurance, not PPP-adjusted (euros)
gbusv_e:	Value of own business (not the owned share thereof), not PPP-adjusted (euros)
sbusval:	Owned share of own business (percentage points)
ownbv_e:	Value of the owned share of own business, not PPP-adjusted (euros)
carv_e:	Value of cars, not PPP-adjusted (euros)
liabv_e:	Value of financial liabilities, not PPP-adjusted (euros)
riskv_e:	Value of risky financial assets, not PPP-adjusted (euros)
homev_p:	Value of the house, PPP-adjusted (euros)
mortv_p:	Value of the mortgage, PPP-adjusted (euros)
oresv_p:	Value of other real estate, PPP-adjusted (euros)
baccv_p:	Value of bank accounts, PPP-adjusted (euros)
bondv_p:	Value of bonds, PPP-adjusted (euros)
stocv_p:	Value of stocks, PPP-adjusted (euros)
mutfv_p:	Value of mutual funds, PPP-adjusted (euros)
irav_p:	Value of individual retirement accounts, PPP-adjusted (euros)
contv_p:	Value of contractual savings for housing, PPP-adjusted (euros)
linsv_p:	Value of life insurance, PPP-adjusted (euros)
gbusv_p:	Value of own business (not the owned share thereof), PPP-adjusted (euros)
ownbv_p:	Value of the owned share of own business, PPP-adjusted (euros)
carv_p:	Value of cars, PPP-adjusted (euros)
liabv_p:	Value of financial liabilities, PPP-adjusted (euros)
riskv_p:	Value of risky financial assets, PPP-adjusted (euros)
homev_f:	Flag for the value of the house
mortv_f:	Flag for the value of the mortgage
oresv_f:	Flag for the value of other real estate
baccv_f:	Flag for the value of bank accounts
bondv_f:	Flag for the value of bonds

stocv_f:	Flag for the value of stocks
mutfv_f:	Flag for the value of mutual funds
irav_f:	Flag for the value of individual retirement accounts
contv_f:	Flag for the value of contractual savings for housing
linsv_f:	Flag for the value of life insurance
gbusv_f:	Flag for the value of own business (not the owned share thereof)
sbusv_f:	Flag for the owned share of own business
carv_f:	Flag for the value of cars
liabv_f:	Flag for the value of financial liabilities
riskv_f:	Flag for the value of risky financial assets

c. Household-level variables

hhomev_e:	Value of the house, not PPP-adjusted (euros)
hmortv_e:	Value of the mortgage, not PPP-adjusted (euros)
hosesv_e:	Value of other real estate, not PPP-adjusted (euros)
hbaccv_e:	Value of bank accounts, not PPP-adjusted (euros)
hbondv_e:	Value of bonds, not PPP-adjusted (euros)
hstocv_e:	Value of stocks, not PPP-adjusted (euros)
hmutfv_e:	Value of mutual funds, not PPP-adjusted (euros)
hirav_e:	Value of individual retirement accounts, not PPP-adjusted (euros)
hcontv_e:	Value of contractual savings for housing, not PPP-adjusted (euros)
hlinsv_e:	Value of life insurance, not PPP-adjusted (euros)
hownbv_e:	Value of the owned share of own business, not PPP-adjusted (euros)
hcarv_e:	Value of cars, not PPP-adjusted (euros)
hliabv_e:	Value of financial liabilities, not PPP-adjusted (euros)
hrav_e:	Value of real assets, not PPP-adjusted (euros)
hgfinv_e:	Value of gross financial assets, not PPP-adjusted (euros)
hnfinv_e:	Value of net financial assets, not PPP-adjusted (euros)
hrisk_e:	Value of risky financial assets, not PPP-adjusted (euros)
hnetwv_e:	Net Worth, not PPP-adjusted (euros)

hhomev_p:	Value of the house, PPP-adjusted (euros)
hmortv_p:	Value of the mortgage, PPP-adjusted (euros)
horesv_p:	Value of other real estate, PPP-adjusted (euros)
hbaccv_p:	Value of bank accounts, PPP-adjusted (euros)
hbondv_p:	Value of bonds, PPP-adjusted (euros)
hstocv_p:	Value of stocks, PPP-adjusted (euros)
hmutfv_p:	Value of mutual funds, PPP-adjusted (euros)
hirav_p:	Value of individual retirement accounts, PPP-adjusted (euros)
hcontv_p:	Value of contractual savings for housing, PPP-adjusted (euros)
hlinsv_p:	Value of life insurance, PPP-adjusted (euros)
hownbv_p:	Value of the owned share of own business, PPP-adjusted (euros)
hcarv_p:	Value of cars, PPP-adjusted (euros)
hliabv_p:	Value of financial liabilities, PPP-adjusted (euros)
hrav_p:	Value of real assets, PPP-adjusted (euros)
hgfinv_p:	Value of gross financial assets, PPP-adjusted (euros)
hnfinv_p:	Value of net financial assets, PPP-adjusted (euros)
hriskfv_p:	Value of risky financial assets, PPP-adjusted (euros)
hnetwv_p:	Net Worth, PPP-adjusted (euros)
hhomev_f:	Flag for the value of the house
hmortv_f:	Flag for the value of the mortgage
horesv_f:	Flag for the value of other real estate
hbaccv_f:	Flag for the value of bank accounts
hbondv_f:	Flag for the value of bonds
hstocv_f:	Flag for the value of stocks
hmutfv_f:	Flag for the value of mutual funds
hirav_f:	Flag for the value of individual retirement accounts
hcontv_f:	Flag for the value of contractual savings for housing
hlinsv_f:	Flag for the value of life insurance
hownbv_f:	Flag for the value of the owned share of own business
hcarv_f:	Flag for the value of cars
hliabv_f:	Flag for the value of financial liabilities

hrav_f:	Flag for the value of real assets
hgfinv_f:	Flag for the value of gross financial assets
hnfinv_f:	Flag for the value of net financial assets
hriskv_f:	Flag for the value of risky financial assets
hnetwv_f:	Flag for the net worth

4. Calculation of Purchasing Power Parities

The PPP adjustment is performed to correct for price level differences across countries. It is performed after all amounts are already expressed in euros, and thus one needs only the relative price levels of the different countries in order to calculate the PPP-adjusted amounts. Data for price levels of the SHARE countries are taken from the OECD (found at <http://www.oecd.org/dataoecd/48/18/18598721.pdf>, dated July 2004). The PPP adjustment is made by dividing the individual country prices by the average of the 11 SHARE countries. This average is computed using as weight the second quarter 2004 nominal private consumption divided by the price level of each country (in order to remove the differential price effect).

It has to be noted that even after adjusting for differences in prices, the values of economic variables are still nominal since they correspond to a basket of goods valued at the same but still current prices. Thus, to compute the PPP-adjusted values one divides the nominal values in euros by the following relative price ratios:

<i>Country</i>	<i>Prices relative to SHARE-11</i>
Austria	0.9918
Belgium	1.0013
Denmark	1.2658
France	1.0296
Germany	1.0296

Greece	0.8501
Italy	0.9446
Netherlands	1.0202
Spain	0.8501
Sweden	1.1241
Switzerland	1.3602

5. Imputation

Imputation is performed using an imputation package in Stata called hotdeck, which is based on the approximate Bayesian bootstrap defined in Rubin and Schenker (1986). This procedure requires classifying the non-missing observations in cells defined by one or more classificatory variables, and from these cells bootstrap samples are drawn. These samples are used to impute the missing observations in each cell. The hot deck is performed for one variable with missing values at a time. In choosing the number of variables to define the cells we face a trade-off. The higher their number is, the better the match between the missing and the non-missing observations, but the smaller the number of observations with non-missing values within the cell. We use multiple imputation during which the hot deck procedure creates five different values for each missing one. This is done by drawing five samples with replacement from the cells of non-missing observations. Further details are given in Section 5.4.

5.1 Imputation of Ownership Variables

Each question about the amount of an item is preceded by a corresponding question about whether this item is owned or not. The ownership questions corresponding to each asset are:

- i) Primary residence – HO002_
- ii) Mortgage – HO013_
- iii) Other Real Estate – HO026_

- iv) Bank accounts, bonds, stocks, mutual funds, respondent's individual retirement account, contractual savings for housing, life insurance – AS002_1, AS002_2, AS002_3, AS002_4, AS002_5, AS002_6, AS002_7, AS002_8
- v) Individual retirement account of the respondent and his/her spouse: AS020_
- vi) Own business – AS041_
- vii) Cars – AS049_
- viii) Financial Liabilities – AS054_1, AS054_2, AS054_3, AS054_4, AS054_5, AS054_6, AS054_7, AS054_8

If an individual gives a response of don't know or refuses to answer the ownership question, then ownership is imputed. In addition there are households in which no individual gives any response for the housing (question HO002_), financial assets (question AS001_) or financial liabilities (question AS053_) section. In that case ownership is imputed for the designated household head. The imputation is done using country and age as classificatory variables for the hot deck procedure.

5.2 Imputation of Amount Variables

Once the ownership question has an original or imputed positive value, the amount is imputed in the following cases:

- i) When the ownership is imputed and the result is positive (flag variable equals 7).
- ii) When the individual gives a response of don't know/refusal and either does not start the unfolding brackets procedure (flag variable equals 5), or does not complete it (flag variable equals 3), or completes it without giving a specific amount as an approximate answer (flag variable equals 2, which is however the value also if the approximate amount is given during the unfolding bracket procedure).

- iii) When the original answer is an illegitimate negative value, a zero while the ownership answer is positive, an implausibly low positive value, a wrong currency answer or a very high outlier (flag variable equals 10).

In the end we divided the variables into three groups according to the criteria by which the cell classification for imputation was made (all imputations were made separately for each country):

- i) *Housing, bank accounts and cars.* These variables contained numerous positive non-missing values, reflecting the wide ownership of the corresponding assets. In the case in which we did not know the bracket value we used age as an additional variable. When we knew the bracket value, we used it together with age.
- ii) *Mortgage.* We needed to link the value of the mortgage to the value of the underlying house, in order to avoid as much as possible the case where the imputed value of the mortgage was greater than the value of the house. Thus, when we did not know the bracket value of the mortgage, we used the bracket value of the house as a classificatory variable; when we knew the bracket value of the mortgage we used it for the imputation. We left out the bracket value of the house because its inclusion would have made the cells too thin.
- iii) *Other real estate, bonds, stocks, mutual funds, individual retirement accounts, contractual savings for housing, life insurance, own business and owned share thereof and financial liabilities.* These variables exhibited relatively few positive non-missing values. We used age to define the imputation cells when we did not know the bracket value, while we used the bracket value for their definition when we knew it.

Following convention, we use a male as the household head, provided his record is in the first two observations of a given household, since typically these are the lines where members of a couple or primary respondents are listed. If there's no male listed in the first two observations, we pick the first female listed as head. Having designated the household head, we had to decide whether to use the individual's or the household head's information (e.g.

age) in order to classify each missing value into cells. Using the individual's characteristics assumes that s/he plays the most significant part in determining the value of (a potentially household-level) variable. On the other hand, the head's information can be more useful in cases where the head does not respond and the answer is provided by someone else purely for convenience reasons. If the household head responds, then each individual has his/her missing values imputed using his/her information. If the head does not respond then the first respondent with missing values is assigned the head's information, while any further respondents' answers are imputed using their own information.

5.3 Imputation of Indirect Stockholding

As already mentioned in Section 2.1 we need to determine what part of mutual funds and individual retirement accounts are invested in stocks, and to this effect we use variables AS019_, AS023_ and AS026_. When these two variables have missing values we impute them using hot deck by country and age.

5.4 Multiple Imputation

We generate five values for each missing one by running the same program five different times using a different seed to perform the hot deck imputation in each run of the program. Thus we generate five different implicate datasets which have identical values when these were not originally missing and potentially five different values for the missing cases. The five datasets are indexed by the variable `implicat` which takes the value 1 for the first dataset, 2 for the second and so on.

It is fundamental to always take into account the fact that we have five different datasets when performing any kind of analysis. This means that one should not use just one of the five datasets nor one should concatenate all five and treat them as one. Rather, one should perform the analysis on each dataset separately and then combine the results from all five datasets using the results of Rubin (1987) [see also Little and Rubin (2002) for a recent survey].

Let $m=1, \dots, M$ index the imputation run (with M in our case equal to 5) and let $\hat{\beta}_m$ be our estimate of interest (e.g. sample median, regression coefficient etc.) from the m^{th} implicate dataset. Then the estimate using all M implicate datasets is just the average of the M separate estimates, i.e.

$$\bar{\beta}_M = \sum_{m=1}^M \hat{\beta}_m$$

The variance of this estimate consists of two parts. Let V_m be the variance estimated from the m^{th} imPLICATE dataset. Then the first magnitude one needs to compute is the average of all M variances, which constitutes the within-imputation variance, i.e.

$$WV_M = \frac{1}{M} \sum_{m=1}^M V_m$$

The second magnitude one needs to compute is the between-imputation variance, which is given by:

$$BV_M = \frac{1}{M-1} \sum_{m=1}^M \left(\hat{\beta}_m - \bar{\beta}_M \right)^2$$

Then the total variance of the estimate is equal to:

$$V_M = WV_M + \frac{M+1}{M} BV_M$$

As Little and Rubin (2002) point out, the second term in the above equation represents the share of the total variance due to missing values.

One can perform a usual t-test of significance employing the following formula to compute the degrees of freedom n equal to:

$$n = (M-1) \left(1 + \frac{1}{M+1} \frac{WV_m}{BV_m} \right)^2$$

The package `hotdeck` in Stata has an option that allows the user to execute many commands using the generated imPLICATE datasets and to combine the results according to the aforementioned rules. In addition, there are 2 other Stata packages, `st0042` and `st0067`, which

can be downloaded from the Internet and perform regression-based analysis using multiple imputation. Furthermore, we provide two additional Stata programs that show how multiple imputation calculations are done and that can be easily modified in order to calculate additional magnitudes of interest:

- a) `mi-trial-descr.do`: calculates means and medians
- b) `mi-trial-ols.do`: performs an OLS regression and calculates coefficients, standard errors, t-statistics, p-values, adjusted R squared, F-test, rmse, log likelihood and likelihood ratio test

Multiple imputation can also be performed in SAS using PROC MI and PROC MIANALYZE, and also in R and S-plus. SPSS version 12 has some problems with performing missing data analysis, as documented in von Hippel (2004). More information on multiple imputation can be found in the following sites:

- i) <http://www.stat.psu.edu/~jls/misoftwa.html>
- ii) www.multiple-imputation.com
- iii) http://www.herc.research.med.va.gov/FAQ_I9.htm

6. Brief Description of the Program that generates the Asset Variables

The program is called `Assets.do`³ and is programmed in Stata. It needs one additional package to run, called `hotdeck`, which can be downloaded from the Internet by executing the command “`net install hotdeck`”. The program broadly consists of the following parts:

- i) The data files are read. These are the three section files, `_cm`, `_ho` and `_as` for all countries (except Belgium for which data are not presently available)
- ii) Head status is determined

³ This program, together with the two aforementioned programs `mi-trial.descr.do` and `mi-trial-ols.do` can be obtained from the corresponding author on an “as-is” basis upon request. They are distributed in the hope that they will be useful, but without warranties of any kind. All original material is provided under a Creative Commons Attribution-ShareAlike license.

- iii) The ownership and amount variables are defined, and the missing observations are recorded. Flag variables are also defined. All these calculations are done using the criteria described in Section 2.
- iv) Imputation of ownership (see Section 5).
- v) Imputation of amount (see Section 5).
- vi) Imputation of indirect stockholding (see Section 5).
- vii) Definition of household-level variables and of aggregates (see Section 2).

References

- Little, R.J.A. and D.B. Rubin (2002), "Statistical Analysis with Missing Data", New York: Wiley, 2nd edition
- Rubin, D.B. (1987), "Multiple Imputation for Nonresponse in Surveys", New York: Wiley
- Rubin, D.B. and N. Schenker (1986), "Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Non-response", *Journal of the American Statistical Association*, 81, 366-74
- von Hippel, Paul T. (2004), Biases in SPSS 12.0 Missing Value Analysis, *The American Statistician*, 58, 160-64